APCC
APEC CLIMATE CENTER

TECHNICAL REPORT

APCC
APEC CLIMATE CENTER

PREFACE

It is our pleasure to present to you the APEC Climate Center (APCC)'s Technical Report 2012, which reports the core outcomes of our research activities from the past year.

Since 2005, APCC, as a hub of climate information in the Asia-Pacific region, has strived to share our analysis and prediction of abnormal climate and to apply this information to regional development. The Center has established the most extensive Multi-Model Ensemble (MME) system for seasonal prediction in the world through its international science network and has provided value-added products to various stakeholders. Recently, APCC has expanded its mandate to include enhancing the capacity of APEC member economies to respond effectively to climate change and variability through better application of climate information.

In 2012, APCC continued to make an effort to improve the quality and quantity of our short-term climate forecasts and our online climate information systems, as information dissemination tools. Additionally, APCC began its endeavor to produce more applicable climate information through interdisciplinary research among various sectors, such as agriculture and hydrology. The following technical report provides more information about our research outcomes from 2012.

In 2013, following APCC's goal to enhance socioeconomic well-being through better utilization of climate information, APCC will continue to improve the quality and accuracy of its climate information, recognizing that the utility of this information is only as good as its quality. We would like to make the best use of our research outcomes in various scientific and application areas. We welcome any feedback on this report or on our services.

My best and warmest regards to all of you.

Dr. Chin-Seung Chung

Director/APEC Climate Center

CONTENTS

APCC
TECHNICAL REPORT 2012-01

Assessment of APCC Multi-Model Ensemble Predictions in Seasonal Climate Forecasting: Retrospective (1983–2003) and Real-time Forecasts (2008–2011)

Dr. Young-Mi Min

ABSTRACT

Since its inception in 2005, the Asia-Pacific Economic Cooperation (APEC) Climate Center (APCC) has devoted considerable effort to developing a multi-model ensemble (MME) prediction system for producing improved and well-validated seasonal global forecasts. Since 2007, the APCC has issued monthly MME seasonal climate predictions for the upcoming three-month period with a one-month lead time, and has disseminated it to APEC member economies. This paper gives a comprehensive documentation of the current status of the APCC operational multi-model performance, with a large set of predictions of global temperature and precipitation in different seasons for the retrospective (1981–2003) and real-time (2008–2011) forecast periods. It focuses on a comparison between the different multi-model combination schemes, in terms of the skill of temperature and precipitation forecasts. In particular, simple averaged MME in which model forecasts are combined with equal weights, two empirically-weighted MMEs based on multiple regression and a calibrated MME based on a step-wise pattern projection method are considered. Results indicate that when considering all aspects of the predictions, (i.e. with different variables, regions and seasons), the methods of the multiple regression-based weighted MMEs consistently show a lower skill than the simple averaged MME. The calibrated MME predictions by model correction provide the most effective way of reducing errors and improving forecasts for both variables than other MME method during the cross-validated 21-year retrospective, as well as during independent 4-year real-time forecast periods. Basics for the rationality of the MME and the possible causes of the failure and success of the different MME methods implemented in the APCC operations, are also discussed.

1. INTRODUCTION

Since its inception in 2005, The Asia-Pacific Economic Cooperation (APEC) Climate Center (APCC) has facilitated the sharing of high-cost climate data and information and promoted capacity-building in order to meet the growing societal and economic interests in the monitoring and prediction of seasonal climate variability and to minimize economic and human losses due to natural disasters. Since 2007, the APCC has issued monthly global temperature and precipitation forecasts for the upcoming three-month period, which are disseminated to APEC member economies via the website (http://www.apcc21.org)

These forecasts utilize deterministic (based on ensemble mean) and probabilistic (based on the full set of ensemble members) interpretations of the well-validated multi-model ensemble (MME) seasonal prediction system. Recently, the APCC has also initiated the sea surface temperature (SST) and El Niño-Southern Oscillation (ENSO) forecasts, based on the MME predictions, using state-of-the-art coupled general circulation models.[1] Currently, 17 prominent operational climate centers and research institutes from 9 APEC member economies participate in the APCC operational MME prediction by routinely providing their predictions in the form of ensembles of global forecast fields (Table 1). The MME technique was designed at around the start of this century to quantify uncertainties in forecasts associated with model formulation (Krishnamurti *et al.* 1999, 2000; Doblas-Reyes *et al.* 2000; Palmer *et al.* 2000; Shukla *et al.* 2000), and has been considered an effective means of improving weather and climate forecasts. As a result, various MME prediction systems are currently utilized at several operational centers (e.g., APCC, ECMWF[2], IRI[3], MSC[4], WMOLC[5]) that routinely provide MME seasonal forecasts. Since its inception, the APCC has devoted considerable effort to developing a MME prediction system for producing improved and well-validated seasonal and regional forecasts in both probabilistic and deterministic frameworks, for research and operational purposes (e.g., Kug *et al.* 2008b; Kang *et al.* 2009; Min *et al.* 2009; Lee *et al.* 2011; Min *et al.* 2011; Sohn *et al.* 2012; Min *et al.* 2012). The operational probabilistic forecasts at the APCC are based on the MME, with the model weights being inversely proportional to the random errors in the forecast probability (Min *et al.* 2009). The probabilistic MME prediction system has also been operationally implemented at the WMOLC since 2011[6] Detailed methodology of the APCC operational probabilistic MME prediction system and its general performance are described by Min *et al.* (2009) based on retrospective and real-time forecasts. Currently, four deterministic MME methods, based on the ensemble means of participating models, are operationally exploited

1) Information available at http://www.apcc21.org/eng/service/fore/lmon/japcc030101_lst.jsp
2) The European Centre for Medium-range Weather Forecasts, UK
3) The International Research Institute for Climate and Society, USA
4) Meteorological Service of Canada, Canada
5) World Meteorological Organization Lead Center, Korea
6) information available at http://www.wmolc.org/multi_model/pMME.php

for seasonal forecasts at the APCC. The first method is a simple averaged MME, in which contributions from all the participating models are equally weighted (i.e., a simple composite method; "SCM"). The second is a calibrated MME obtained from the adjusted (or corrected) single-model predictions, based on a stepwise pattern projection method ("SPM"; Kug *et al.* 2008b).

Preliminary assessments of the skill of the APCC deterministic methods were provided in the APCC technical report by Kryjov *et al.* (2006). However, many models that comprise the APCC model set have been upgraded since that time and an essential number of new assessments, characterizing the skill of the individual models and different deterministic multi-model combinations, has since been obtained. This paper provides an updated and comprehensive assessment of the APCC operational deterministic MME forecasts based on a large set of predictions, focusing particularly on comparisons between the different MME schemes (i.e., SCM, SPM, MRG, and SSE).

A number of studies have focused on the comparison between the different methods used with a combination of individual model forecasts (e.g., Kharin and Zwiers, 2002; Yun *et al.* 2003, 2005; Yoo and Kang, 2005; DelSole, 2007). However, alongside the method of combining forecasts, the skill of the multi-model combination depends upon the regions and periods of prediction of the participating models. Nowadays, the APCC operates the largest multi-model ensemble prediction system and it has obtained essential experience in the development and assessment of multi-model predictions. This paper is, therefore, intent on providing a comprehensive coverage of the various aspects of multi-model predictions, based on both historical and real-time forecasts. In order to achieve this, a large set of predictions obtained from the (multiple regression-based) weighted (MRG and SSE) and calibrated MME prediction systems (SPM) are assessed for the prediction of a 12 running 3-month mean of 850-hPa temperature and precipitation, one month ahead, in comparison with the (non-weighted) simple averaged MME prediction system (SCM), during the retrospective forecast period of 1983–2003 and during the real-time period of 2008–2011. With the increasing popularity of multi-model forecasting, it is, however, important to understand the virtues and limitations of the MME. In this regard, a basis for the rationality of the MME, with an explanation for the possible causes of failure and success of the different MME schemes used in this study, will be also discussed.

The paper is organized as follows: Section 2 describes the participating models in the APCC MME prediction, their retrospective and real-time forecast datasets and corresponding observations, along with a brief explanation of the verification metrics. Section 3 briefly introduces the methods of the APCC deterministic MME prediction systems. In Section 4, we evaluate the overall performance of the participating single-models and their simple multi-model predictions. Sections 5 and 6 present the assessment of the multi-model predictions obtained from the different MME schemes, with a discussion on the basis for a rationality of the multi-model, together with possible causes of the failure and success of the different MME schemes, respectively. Finally, Section 7 contains a brief summary of the results. A brief summary of the results follows in Section 7.

2. DATA AND FORECAST SKILL MEASURES

The models examined consist of the 12 dynamical climate prediction systems from APEC member economies that currently participate in the APCC operational one-month lead, three-month mean, MME forecast. Table 2 gives a brief summary of the models' specifications for the seven two-tier (CWB, GCPS, GDAPS, AGCM2, AGCM3, SEF, and NIMR) and five one-tier systems (BCC, JMA, NCEP, PNU, and POAMA). The models show a large range of resolutions and ensemble sizes, and their retrospective forecast datasets match the requirements of the Seasonal Prediction Model Intercomparison Project/Historical Forecast Project (SMIP/HFP), or the Coupled Model Intercomparison Project (CMIP). All models have generated ensemble retrospective forecasts for the common period 1983–2003 and real-time forecasts for the common period 2008–2011. In the present study, we focus on one-month lead, three-month mean, forecasts of temperature at 850 hPa (hereinafter, "temperature") and precipitation, every month. Note that the participating models in the APCC MME prediction among 12 climate models are slightly different from for each target season due to the operational situations at that time. The models' datasets are interpolated to a common resolution of 2.5°lon × 2.5°lat grid, similar to that of the observed data. In the study, we used the models' anomalies, by removing the climatological model biases. For each individual model, the anomalies are estimated as departures from their climatology over the training period using a one-year out, cross-validation (Wilks 1995). The same procedure for the estimation of anomalies is applied to the observed dataset.

The data used for verification of the retrospective and real-time forecasts of temperature and geopotential height were obtained from the NCEP-Department of Energy (DOE) reanalysis 2 data (Kanamitsu *et al.* 2002). The observed precipitation data used in this study are the Climate Anomaly Monitoring System and Outgoing longwave radiation Precipitation Index data (CAMS OPI; Janowiak and Xie 1999). The CAMS OPI is a precipitation analysis created by merging ground-based rain gauge observations with satellite rainfall estimates to obtain real-time monthly analyses of global precipitation. We used the Niño 3.4 index based on optimum interpolation (OI) version 2, monthly mean SST (Reynolds *et al.* 2002), obtained from the Climate Diagnostics Center (CDC) of the National Oceanographic and Atmospheric Administration[7] (NOAA). The metrics used to measure the forecast skill of the APCC single-model and multi-model predictions include the temporal correlation coefficient, the anomaly pattern correlation coefficient, and the root mean square error (RMSE). The Student's t-statistic (t-test) was used to assess the statistical significance of the estimated temporal correlation at the 5% level. Following recommendations from the WMO Standardized Verification System for Long-Range Forecast (SVS-LRF; WMO 2002), the verification scores were also produced over several sub-regions to estimate large-scale verification statistics, in order to evaluate the overall skill of the forecast system, including the globe, tropics (20°S–20°N), and northern (>20°N) and southern hemispheres (>20°S). In addition, to provide a regionalized assessment of the forecast system, the Asian-Australian monsoon regions; including the Indian Monsoon (60–105°E, 5–30°N), Western North Pacific Monsoon (105–160°E, 5–20°N), East Asian Monsoon (110–140°E, 20–45°N), and Australian Monsoon (105–160°E, 20–5°S), are

7) data available at http://www.cpc.ncep.noaa.gov/data/indices/

used in the study. Note that for the 21-year retrospective forecast (1983–2003), the one-year out cross-validation method was used for the calculation of skill measures, using the 20-year training period with the target year withheld. For the independent real-time forecasts (2008–2011), we used the whole 21-year hindcast data as the training period for each real-time forecast.

Table 1 Acronyms for names of institutes and their models used in the text.

Country	Organization/Institute
Australia	Australian Bureau of Meteorology (BoM)
Canada	Meteorological Service of Canada(MSC)
China	Beijing ClimateCenter (BCC)
	Institute of Atmospheric Physics of China(IAP)
Japan	Japan Meteorological Agency (JMA)
Korea	Korea Meteorological Administration (KMA)
	National Institute of Meteorological Research of Korea(NIMR)
	Seoul NationalUniversity (SNU)
	Pusan NationalUniversity (PNU)
Peru	Meteorological and Hydrological Weather Service of Peru(SENAMHI)
Russia	Main Geophysical Observatory of Russia(MGO)
	Hydrometeorological Centre of Russia(HMC)
Chinese Taipei	Central Weather Bureau of Chinese Taipei(CWB)
USA	Center for Ocean-Land-Atmosphere Studies (COLA)
	International Research Institute for Climate and Society (IRI)
	National Aeronautics and Space Administration (NASA)
	National Center for Environmental Prediction (NCEP)

Table 2 Description of 10 dynamical seasonal prediction models used in the study.

Model Acronym	Institution (Country)	Resolution	Ensemble Size (H/F)	SST Specification (H/F)	Reference
BCC	BCC (China)	T63L16	8/8	Predicted SST/ Predicted SST	Ding *et al.* (2000)
CWB	CWB (Chinese Taipei)	T42L18	10/10	Predicted SST/ Predicted SST	Liou *et al.* (1997)
GCPS[8]	SNU (Korea)	T63L21	12/12	Predicted SST/ Predicted SST	Kang *et al.* (2004)
GDAPS[9]	KMA (Korea)	T106L21	20/20	Predicted SST/ Predicted SST	Park *et al.* (2002)
JMA	JMA (Japan)	T95L40	5/51	Predicted SST/ Predicted SST	Takaya *et al.* (2010)
AGCM2	MSC (Canada)	T32L10	10/10	Persistent ERA[10]-40-SST/ Persistent CMC[11] SST	McFarlane *et al.* (1992)
AGCM3	MSC (Canada)	T63L32	10/10	Persistent ERA40-SST/ Persistent CMC SST	Scinocca *et al.* (2008)
SEF[12]	MSC (Canada)	T95L27	10/10	Persistent ERA40-SST/ Persistent CMC SST	Ritchie (1991)
NIMR	NIMR (Korea)	5ox4oL17	10/10	Persistent[13] OISST/ Persistent OISST	Back *et al.* (2002)
NCEP	NCEP (USA)	T62L64	15/15	Predicted SST/ Predicted SST	Saha *et al.* (2006)
PNU	PNU (Korea)	T42L18	5/10	Predicted SST/ Predicted SST	Park *et al.* (2004)
POAMA[14]	BMRC (Australia)	T47L17	10/10	Predicted SST/ Predicted SST	Zhong *et al.* (2005)

8) Global Climate Prediction System
9) Global Data Assimilation and Prediction System
10) ECMWF 40 year Re-Analysis
11) Canadian Meteorological Centre
12) Spectral aux éléments finis Model
13) National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation SST
14) Predictive Ocean-Atmosphere Model for Australia

3. METHODOLOGY OF THE APCC MME PREDICTION SYSTEM

As aforementioned, the APCC has been operationally implementing four deterministic MME prediction systems; SCM, MRG, SSE, and SPM, for one-month lead, three-month mean temperature and precipitation predictions. A basic description of each scheme will be separately introduced in the following sections, separately.

3.1 Simple composite method

Current seasonal climate models, including state-of-the-art coupled models, persist in containing systematic and random errors that degrade seasonal climate predictions, as mentioned in many previous studies(e.g., Feddersen *et al.* 1999; Kang *et al.* 2004; Wang *et al.* 2008; Jin *et al.* 2008; Kug *et al.* 2008b; Lee *et al.* 2010, 2011b). Errors of the anomaly component are related to the climate model's incorrect performance when simulating anomalies. Such anomalies tend to be misplaced when contrasted with the observed dataset(Kang and Shukla 2006). The spatial shifts of the simulations can be corrected by statistical correction methods based on the linear correlation between the model and observed patterns; a so-called "pattern projection technique" (e.g., Kang *et al.* 2004; Feddersen *et al.* 2005; Kang and Shukla 2006).

The most simple and widely used method for constructing a multi-model prediction from single-model ensembles is to use a simple averaged MME, where the contribution of each model is equally weighted ("SCM"; e.g., Doblas-Reyes *et al.* 2000; Peng *et al.* 2002; Palmer *et al.* 2004; Wang *et al.* 2009). The SCM forecast constructed with bias-corrected data is given by

$$Y_t \frac{1}{N}\sum_{i=1}^{N}(F'_{i,j})$$

where, $(F'_{i,j})$ is a forecast anomaly of the i^{th} model at time t, calculated by $F_{i,j}-\overline{F_i}$, with $F_{i,j}$ and $\overline{F_i}$ being the forecast and climatology of the i^{th} model, respectively; and N is the number of individual models involved. Therefore, the results of the SCM are generated by a combination of bias-corrected single-model forecast anomalies.

In this scheme, the ensemble mean assigns the same weight of $1/N$ to each of the N member models at all the grid points, regardless of their relative performance.

3.2 Point-wise multiple regression method

The second scheme is an empirically weighted MME with coefficients computed using multiple regression ("MRG"; Krishnamurti *et al.* 2000; Yun *et al.* 2003). The conventional multi-model prediction, constructed with bias-corrected data is given by

$$Y_t = \sum_{i=1}^{N} a_{i,t} (F'_{i,t})$$

where, $a_{i,t}$ is a regression coefficient of the i^{th} model at time t, obtained by a mean squared error minimization procedure. That is, for each grid point, the respective weightings of the single-model are generated using a point-wise multiple regression technique based on the training period. To compute the regression coefficients for a set of different model forecasts in this study, a singular value decomposition (SVD) is applied. Note that the regression coefficients are calculated in the one-year out, cross-validation mode.

3.3 Synthetic super-ensemble method

The super-ensemble scheme is the empirically weighted MME in which the individual model forecasts are statistically combined during the training period using multiple regression, with the skill of each single-model implicitly factored into the super-ensemble (or MME) forecast (Yun *et al.* 2005). The skill of the super-ensemble relies strongly on the past performance of the individual models used in its construction. The algorithm involves an empirical orthogonal function (EOF) filtering of the individual model datasets (i.e., actual datasets), prior to the construction of the MME. That is, this technique generates a new dataset using the individual models by finding a consistent spatial pattern between the individual model forecasts and observations. MRG is then applied to these EOF-filtered data. Therefore, the newly generated EOF-filtered data set is used as input for the construction of the MME as follows:

$$Y_t = \sum_{i=1}^{N} a_{i,t}\,(\widehat{F_{i,t}})$$

where $\widehat{F_{i,t}}$ is the EOF-filtered forecast anomaly of i^{th} model at time t.

3.4 Step-wise pattern projection method

The last scheme, SPM, is a calibrated MME obtained from the adjusted (or corrected) single-model predictions, based on a step-wise pattern projection method (Kug *et al.* 2008b). The main idea of this method is to produce a prediction of the predictand, by projecting the spatial pattern of the predictor field onto the covariance pattern between the large-scale predictor field and the one-point predictand (Lee 2003).

The model equation is as follows:

$$Y(t) = \alpha \cdot X(t)$$

$$\alpha = \frac{\frac{1}{T}\sum_{t}^{T} Y(t) \cdot X(t)}{\frac{1}{T}\sum_{t}^{T} X^2}$$

$$X(t) = \sum_{x}^{D} Cov(x) \cdot \psi(x,t)$$, and

$$Cov(x) = \frac{1}{T}\sum_{t}^{T} Y(t) \cdot \psi(x,t)$$,

where x and t represent spatial and temporal grid points, respectively. $X(t)$ indicates a time series projected by the covariance pattern between the predictand $Y(t)$ and the predictor field $\psi(x,t)$ in a certain domain D. The covariance pattern indicates a pattern of the model prediction that is related to the observed predictand. The parameter a is a regression coefficient of the projected time series on the predictand during a training period, T. To select the predictor domain, the correlation coefficients are calculated between the predictand and the two-dimensional predictors at each grid point during a training period. On the basis of the selected predictor domain, a statistically corrected prediction is produced by the pattern projection, according to the aforementioned equation. The final MME prediction is obtained as a simple average of the corrected single-model prediction with equal weighting. For more details on the SPM procedures, refer to Kug *et al.* (2008a and 2008b).

4. PERFORMANCE OF APCC SINGLE-MODEL AND MULTI-MODEL PREDICTIONS

Before assessing the four different MME schemes used in the APCC operations, the general performance of the participating single-models in the APCC MME prediction for different seasons in different large-scale regions was investigated, along with their simple averaged MME prediction, in terms of the seasonality, spatial distribution and the interannual variation of the forecast skills.

4.1 Seasonality and spatial distribution of forecast skill

In this section, we first present the overall performance of the APCC single-model predictions for three-month mean temperature and precipitation, with one-month lead and one-month frequency, on the basis of the retrospective forecasts for the period 1983–2003. Figures 1 and 2 show the seasonal cycles of temperature and precipitation forecast skill of the single-model predictions, in terms of temporal correlation, for all 12 running 3-month mean over large-scale regions (i.e., globe, tropics, and northern and southern hemispheres), recommended by WMO SVS-LRF. Here, the estimated scores are first calculated for each grid point and then averaged over the regions. The averaged skill of all the single-model predictions and the skill of their simple averaged MME with equal weighting (SCM) are also presented.

The results from Figs. 1 and 2 revealed three important points. Firstly, it was found that the general performance of the single-model predictions across all seasons has a higher skill for temperature than for precipitation over all large-scale regions.

Relatively high levels of skill are found in the tropics for both variables, as shown in many studies (e.g., Palmer *et al.* 2004; Min *et al.* 2009; Wang *et al.* 2009; Lee and Wang 2012; Min *et al.* 2012; Jia *et al.* 2012). Secondly, it was also clearly demonstrated that the predictions of the SCM give a generally better performance than any single-model prediction across all seasons, regions, and variables. There are only a few cases where the single-model performance is better or comparable to that of the multi-model. However, it should be noted that the main advantage of using a multi-model is not to gain a large improvement in comparison with the respective best single-model in individual cases, but rather to gain a consistently better performance of the multi-model when considering all aspects of the predictions (Doblas-Reyes *et al.* 2005; Hagedorn *et al.* 2005; Min *et al.* 2009; Wang *et al.* 2009; Min *et al.* 2012). The third important point to note in Figs. 1 and 2 is that the forecast skill of the multi-model predictions is consistently better than the averaged skill of all the single-model predictions. The skill of the multi-model prediction is proportional to the averaged skill of the single-model forecasts and inversely proportional to the variance of the multi-model forecasts (Yoo and Kang 2005), the reduction of which, as compared with variances of single model forecasts, being provided by a mutual offset of the single models' errors.

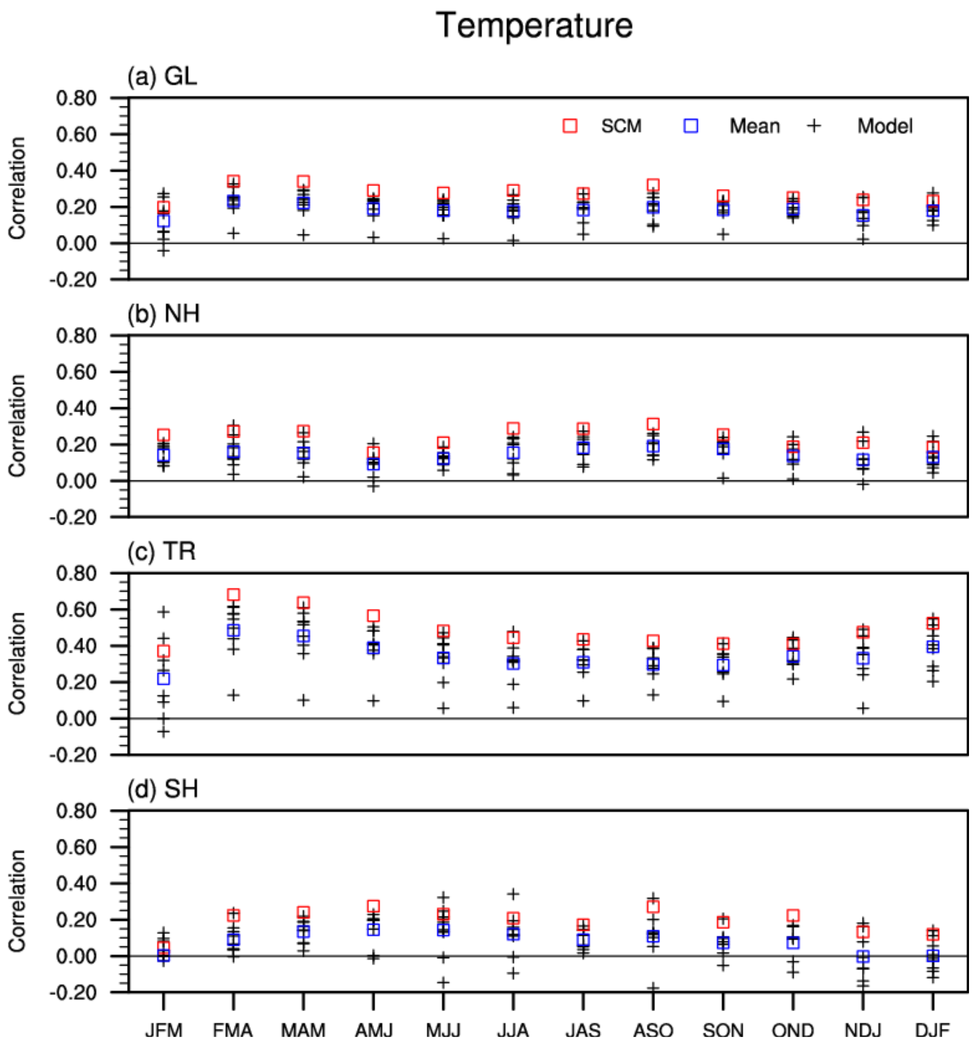

Figure 1 Seasonal distribution of temporal correlations for a single-model (Model; cross), the skill average of the single-model prediction (Mean; blue square), and the simple averaged multi-model prediction with equal weightings (SCM; red square) of temperature for the period 1983–2003 over several regions; globe (GL), northern hemisphere (NH), tropics (TR), and southern hemisphere (SH).

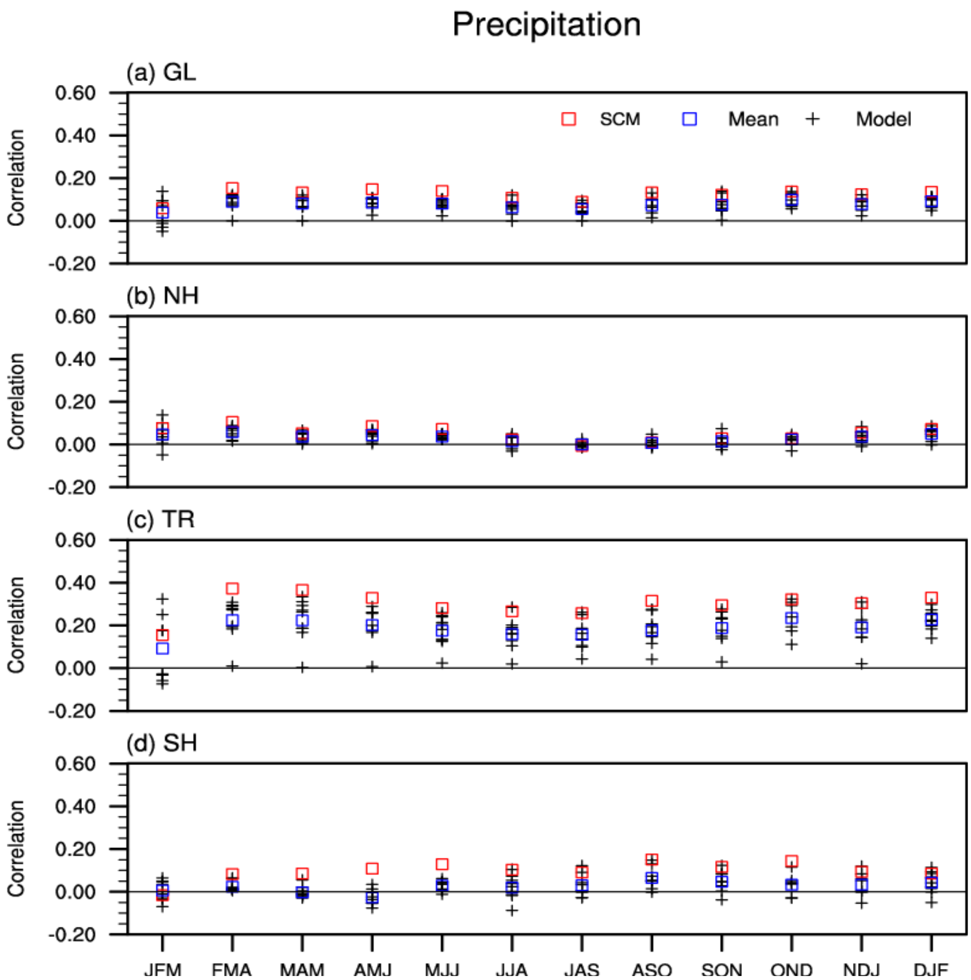

Figure 2. Same as Fig. 1, except for precipitation.

To investigate the spatial distribution of the forecast skills, we evaluated the one-month lead seasonal mean multi-model predictions of temperature and precipitation for the boreal summer (JJA) and winter (DJF) in Fig. 3. Figure 3 also includes significance tests of the estimated temporal correlation at the 5% level, based on the one-tailed Student t-test. In general, the simple averaged MME prediction is more skillful than the single-model predictions for both variables and seasons, with a wider coverage of the areas in which the estimated scores are statistically significant (not shown). As shown in Figs. 1 and 2, temperature is predicted with a higher skill and better coverage than precipitation over the globe, in terms of the temperature correlation. Relatively high levels of skill for temperature prediction are found over most of the tropics and even in the middle latitude regions, such as in Western Europe and Eastern Canada. While skill for precipitation is generally modest, indicating that the estimated area-averaged correlation over the globe for June-July-August (JJA) and December-January-February (DJF) is 0.14 (0.18), the areas where the estimated scores are statistically significant is confined to the tropical Pacific regions. Comparison of the seasons shows that, in DJF, significant skills of both temperature and precipitation are well extended from the Martine continent toward the Indian Ocean, to the extratropics in the northeastern Pacific, and also to North America and Southern Africa. In addition, the most skillful score (> 0.9) is found in the eastern equatorial Pacific, but only in the DJF season. The overall expansion in the skill of the model may be a result of its capacity to capture the ENSO teleconnection. As a result, the area-averaged scores of temperature and precipitation in the boreal winter show the value of 0.32 and 0.18, respectively, which is higher than that in the boreal summer.

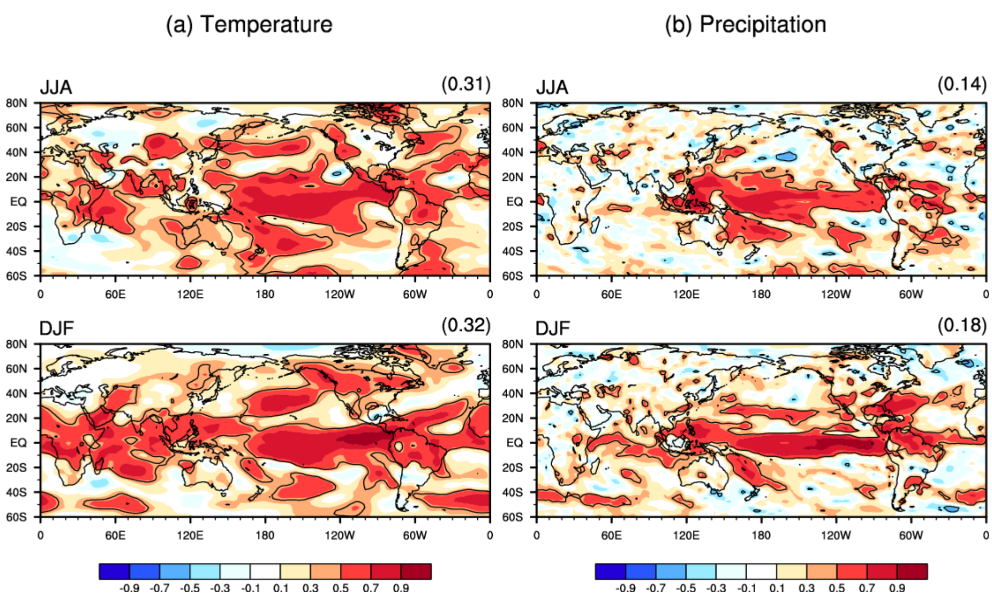

Figure 3 Spatial distribution of temporal correlation for simple averaged multi-model ensemble with equal weighting of (a) temperature and (b) precipitation in JJA and DJF seasons for the period 1983–2003. The area-averaged scores are also displayed in the plot. The contour lines indicate that the estimated score is statistically significant at the 5% level using the one-tailed Student t-test.

4.2 Interannual variation of forecast skill

Figures 4 and 5 show the temporal evolution of the forecast skill of both the single-model and the simple averaged MME predictions for JJA and DJF mean temperature and precipitation, in terms of pattern correlation, over the 21-year period. Similar results can be seen in Figs. 1–3 for different skills measured, illustrating that the SCM predictions generally outperform the single-model predictions for both variables in different seasons. This indicates that the overall skill of the SCM over the globe is consistently positive in most of the years. An interesting feature is that the year-to-year variation in the overall skill over the 21-year period is strongly related to the ENSO variability. The forecast skill of the SCM predictions has a clear relationship with the amplitude of the Niño 3.4 Index(blue dashed line in Figs. 4 and 5), especially in the boreal winter with a correlation value of 0.67 for temperature and 0.82 for precipitation. A relatively lower level of skill for both variables and seasons tends to occur during the transition or during normal ENSO phases. Note that the forecast skills for both variables are higher in the boreal winter than summer, which is likely to be related to the seasonality of the ENSO cycle: El Niño(La Niña) tends to warm (cool) the tropical atmosphere more predictably and strongly near, and following, its mature phase late in the calendar year (Wang *et al.* 2009; Barnston *et al.* 2010). In addition, correlations between the absolute value of the Niño 3.4 index and the forecast skill for precipitation are generally higher than that for temperature, regardless of the season. This finding is in agreement with the results of . (2010) who also demonstrated that the greatest impact of ENSO on the forecast skill of temperature occurs four months following the ENSO peak for both ENSO phases, due to the influence on the forecast skill by a delayed temperature response in both the tropics and extratropics (Kumar and Hoerling, 2003). This was earlier documented in the context of the atmospheric bridge (Lau and Nath 1996; Alexander *et al.* 2002). In the case of precipitation, a simultaneous positive relationship with both phases of ENSO was noted. This supports results from many previous studies showing that the ENSO variability is the sole source of the seasonal precipitation forecast skill (e.g., Wang 2009; Barnston *et al.* 2010).

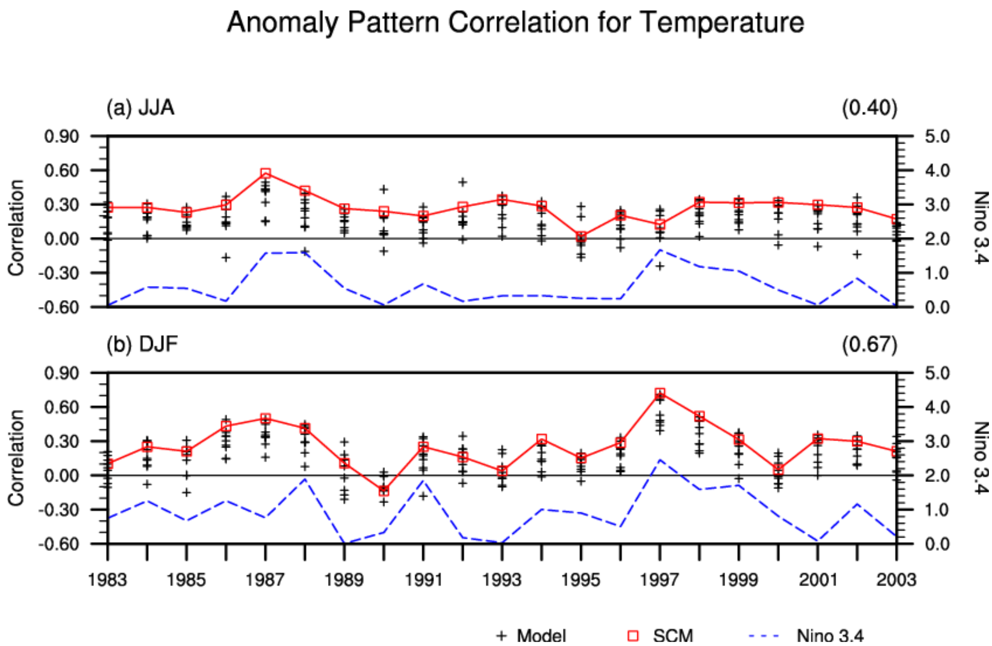

Figure 4 Time series of anomaly pattern correlation for single-model ensembles (Model; cross) and simple averaged multi-model ensemble with equal weightings (SCM; red square) of (a) JJA mean and (b) DJF mean temperature, for the period 1983–2003 over the globe. The blue-dashed line indicates the amplitude of Niño. Correlations between the SCM forecast skill and amplitude of Niño 3.4 Index are also presented in the plot.

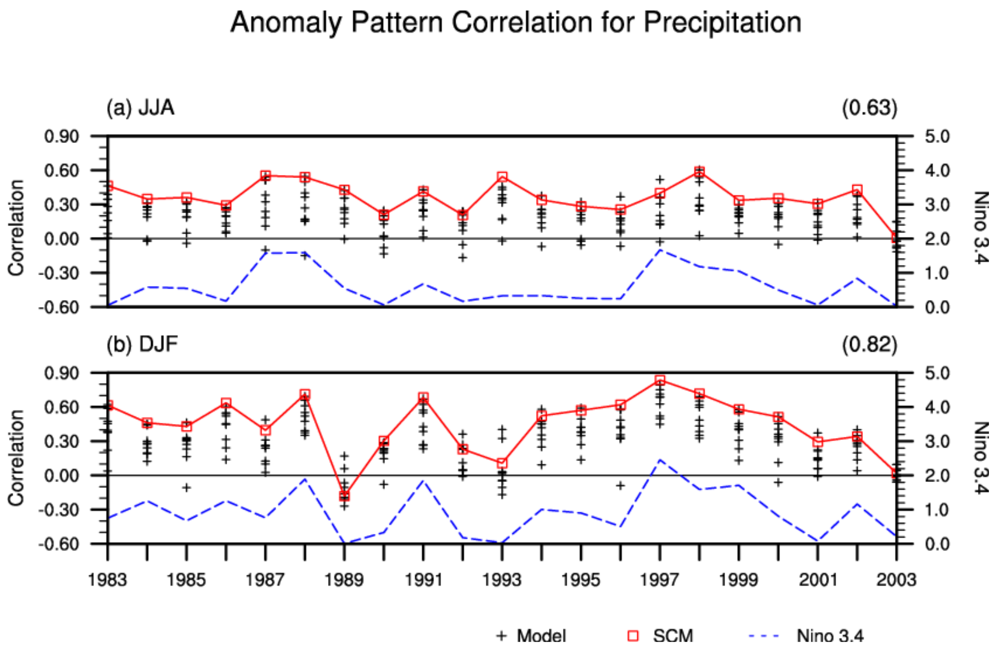

Figure 5 Same as Fig. 4, except for precipitation.

5. COMPARISON OF APCC MULTI-MODEL PREDICTION SYSTEMS

The previous section (including Figs. 1–5) described the overall performance of the APCC single-model predictions and their simple averaged multi-model prediction (SCM) for the period 1983–2003, both spatially and temporally. In this section, we comprehensively assess the forecast skills obtained from the different APCC operational MME prediction systems (SCM, MRG, SSE, and SPM), based on 21-year (1983–2003) retrospective and 4-year (2008–2011) real-time forecasts. As described in Section 3, MRG and SSE are empirically weighted MMEs with coefficients computed using multiple regression and the use of multiple regression with the empirical EOF-filtered dataset, respectively. SPM is a calibrated MME obtained from the corrected single-model predictions, based on a step-wise pattern projection method, and then simply averaging them with equal weighting.

5.1 Retrospective forecast

5.1.1 Global temperature and precipitation

Figures 6 and 7 show the zonal mean temporal correlation as a function of 12 running 3-month mean multi-model predictions for temperature and precipitation, obtained from different MME methods (left panel in Figs. 6 and 7). To explore how the empirically weighted MME methods (using multiple regression; MRG and SSE) and calibrated MME method (using statistical correction; SPM) can improve the performance of the multi-model predictions, their skill improvements are also displayed with respect to the simple MME (SCM; right panel in Figs. 6 and 7), considered as a reference forecast. It was clearly found that the SCM and SPM predictions generally perform better over most latitudinal zones than the MRG and SSE predictions, across all variables and seasons. Their high levels of forecast skills are extended well into the middle- and high-latitudinal zones, especially for temperature, and significant skills are widely found over the tropics throughout all seasons. However, significant skills obtained from (multiple regression-based) weighted MME methods (MRG and SSE) are confined only to the equatorial zone, especially during the boreal cool seasons. Another apparent feature shown in Figs. 6 and 7, particularly visible from a comparison of the SCM and SPM predictions, is the increase in skill when the single-model predictions are corrected by the statistical correction method, SPM. This result is valid for both variables, and illustrates the advantage of model correction using the step-wise pattern projection method. On average, the positive impact of the calibrated MME with its corrected single-model predictions for temperature, is relatively larger in the tropics and southern hemisphere. For precipitation, the model correction contributes to the increase in the forecast skill in most latitudinal zones across all seasons. This result is in agreement with that of Min *et al.* (2012), who only focused on the JJA mean temperature and precipitation.

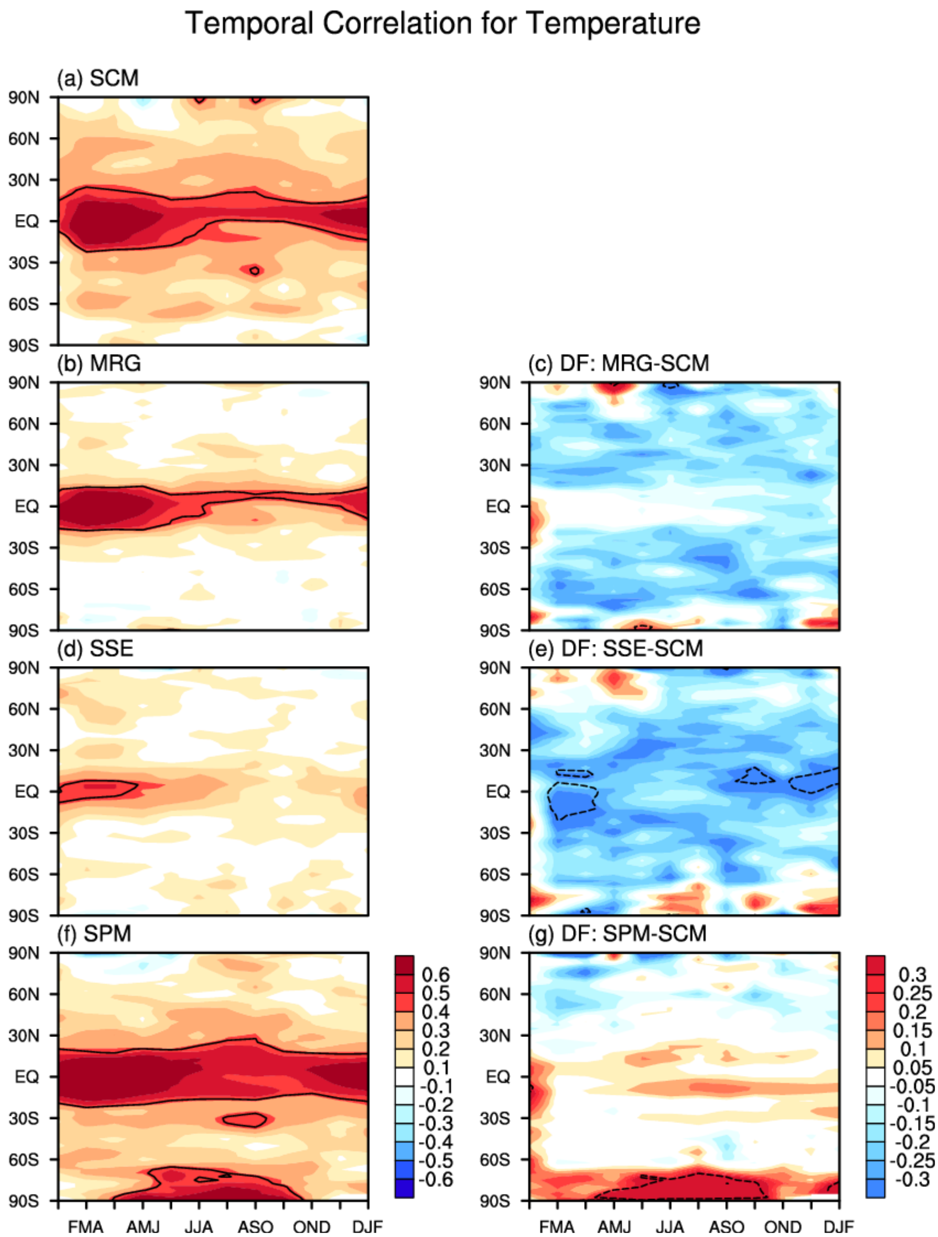

Figure 6 Zonal mean temporal correlation of the SCM, MRG, SSE, and SPM predictions (a, b, e, f) for 3-month mean temperature for the period 1983–2003. The difference maps (DF) show the skill difference of the MRG, SSE, and SPM predictions in relation to SCM (c, e, g), respectively. The solid (dashed) line in the left panels (right panels) represents the estimated score (the difference between the two scores) being statistically significant at the 5% level.

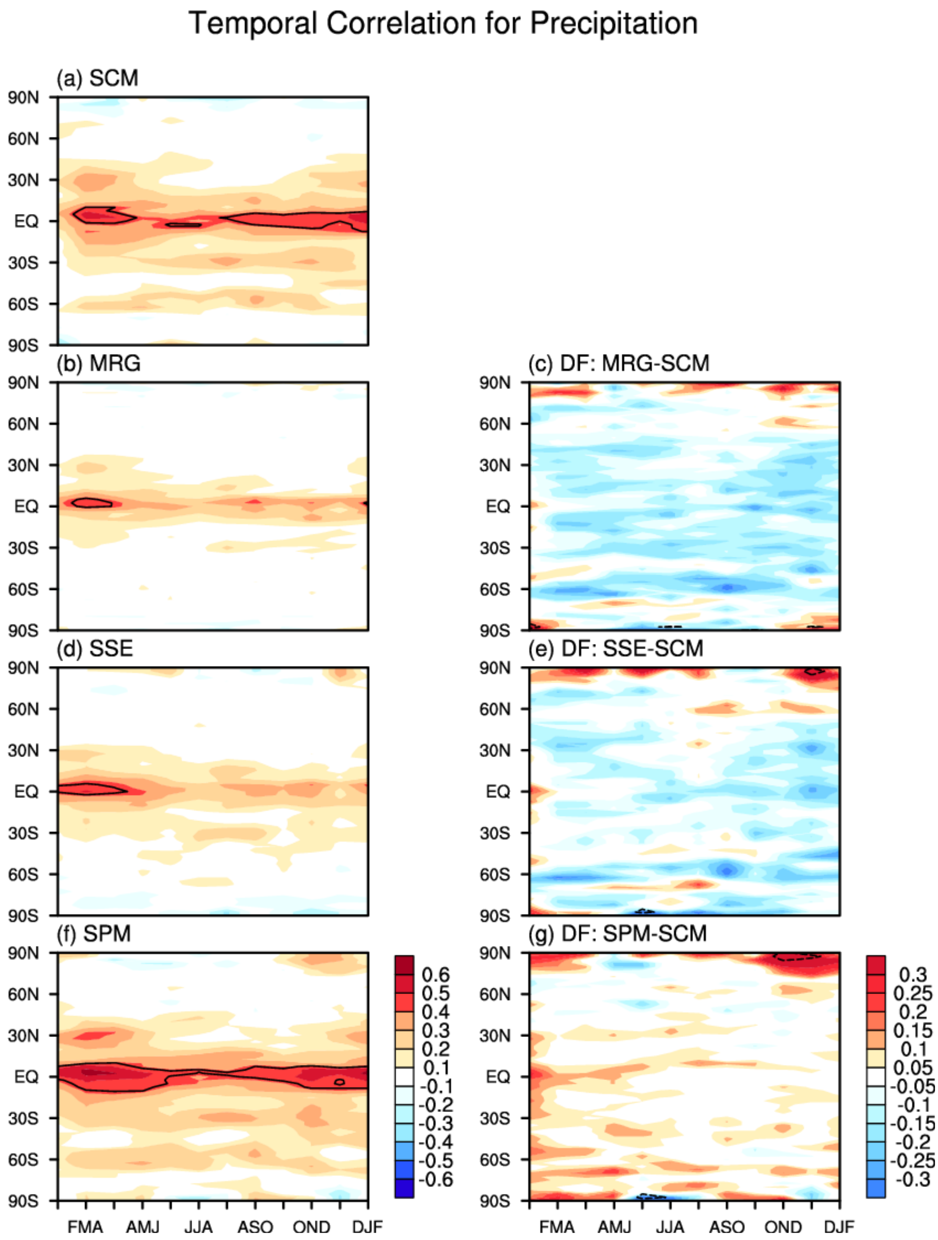

Figure 7 Same as Fig. 6, except for precipitation.

To explore how the model correction based on the step-wise pattern projection method can affect the performance of the multi-model prediction for each variable, the relative improvement of the forecast skill from the simple multi-model prediction to the calibrated multi-model prediction was investigated in Fig. 8. The skill difference between the models was calculated and divided by the skill of the simple multi-model prediction, given a variable and season. That is, a positive value corresponds to the performance of the SPM to a greater extent than that of the SCM, and the degree of the skill increase shows the efficiency of the model correction in the improvement of the multi-model prediction in comparison with the SCM prediction. It is noted that the skill improvement of the SPM, in terms of the area-averaged correlation over the globe, can be found for both variables and for most seasons. The averaged skill of temperature and precipitation across all seasons increases by 12% (with a range of -9% to 33%) and 26% (with a range of 4% to 64%), respectively, from SCM to SPM. The reason for the relatively large skill improvement in precipitation may be due to the low performance of climate models in predicting precipitation in comparison with temperature, which may provide a greater possibility of correcting errors for precipitation than for temperature using a statistical approach. However, further study is necessary to better determine why there is a larger effectiveness in the improved forecast skill for precipitation, rather than for temperature, when using the SPM.

To summarize, it has been demonstrated that the empirically weighted MMEs with coefficients using multiple regression (MRG and SSE) show a consistently lower skill than the reference forecast (SCM), in terms of the seasonal mean global temperature and precipitation predictions for the period 1983–2003. Model correction using the step-wise pattern projection has a positive effect in improving the multi-model predictions of temperature and precipitation, indicating more skillful forecasts than other schemes when considering all aspects of the predictions (e.g., variable, region and season).

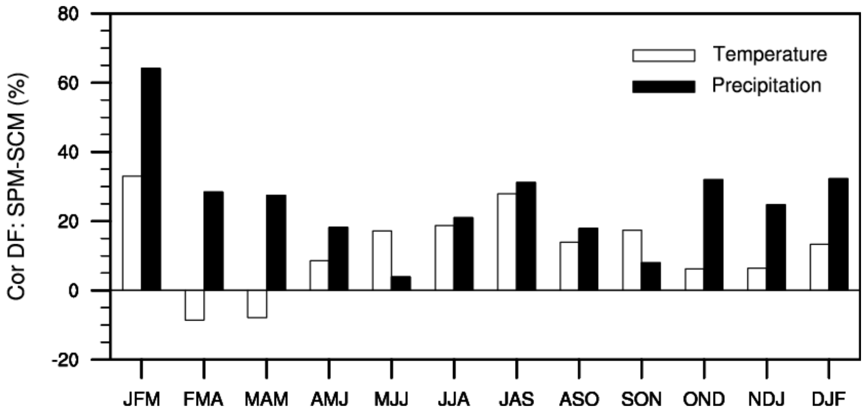

Figure 8 Relative difference of the temporal correlation between the SCM and SPM predictions of 3-month mean temperature (white bar) and precipitation (black bar) over the globe for the period 1983–2003. The relative difference is calculated by the difference between the skills of SCM and SPM, divided by the SCM for each season and variable individually.

5.1.2 Regional monsoon

We used the monsoon domains as one of the metrics to provide a regionalized assessment, to evaluate the capability of the weighted and calibrated MME prediction systems in simulating regional monsoon precipitation, with respect to the simple MME prediction system. We focused on one-month lead three-month mean precipitation over the Asian-Australian monsoon regions, an area which the APCC is particularly interested in; including Indian Monsoon (IM, 5–30°N, 60–105°E), Western North Pacific Monsoon (WNPM, 5–20°N, 105–160°E), East Asian Monsoon (EAM, 20–45°N, 110–140°E) and Australian Monsoon (AM, 20–5°S, 105–160°E) Figure 9 shows the anomaly pattern correlation – RMSE diagram for precipitation for each 12 running 3-month mean and their averages, given a MME method and region. The estimated skills are obtained by computing the anomaly pattern correlation and RMSE for each year and then averaging them.

In general, the overall forecast skill for precipitation prediction over the EAM region, especially in the boreal summer (not distinguishable), is still limited in the present forecasting system, even with the MME predictions, where the estimated forecast skill in terms of anomaly pattern correlation is almost zero. The possible reason for the low skill is that most individual models tend to underestimate the second peak from July to September, during which tropical cyclones account for a significant amount of the precipitation in this region. However, the models are unable to adequately resolve the tropical cyclone rainfall (e.g., Kang *et al.* 1999; Wang and Lin 2002; Lee *et al.* 2010). From our results it can also be seen that the MME predictions have relatively large biases in amplitude over the WNPM region, which may be due to the underestimation of summer rainfall (especially from mid-July to late August) and an overestimation of winter and spring rainfall (as demonstrated by Lee *et al.* 2010).

As discussed in the previous section related to assessments in a global sense, similar results can be found from regionalized assessments. In terms of the RMSE, the performance of multi-model predictions with weighed MME schemes is consistently lower than those of the simple and calibrated MME ones, for both variables and all monsoon regions. The same is true in terms of the anomaly pattern correlation. The performance of the SCM and SPM predictions are generally comparable each other. However, it is further noted that SPM has a marginally better skill than the SCM, with a higher correlation and smaller RMSE across all monsoon regions. This finding implies that the calibrated MME prediction by model correction is effective in reducing errors and improving forecast skills not only for most of the globe (Figs. 6–8), but also for the regional monsoon domains (Fig. 9)

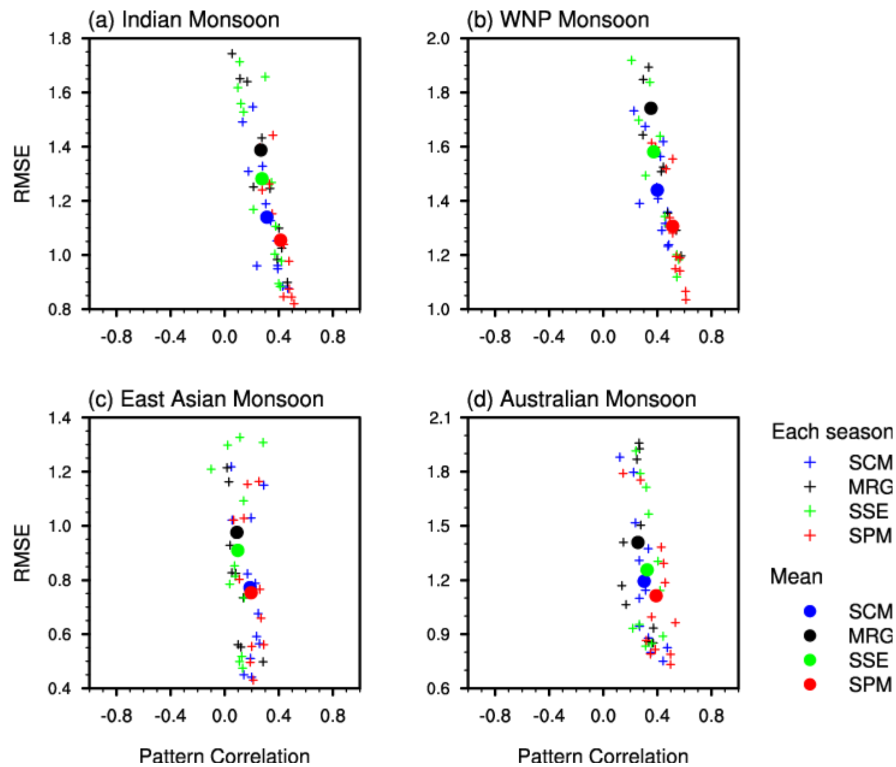

Figure 9 Anomaly pattern correlation – Root mean square error diagram for precipitation prediction obtained from SCM, MRG, SSE, and SPM for each season and their mean: (a) Indian monsoon, (b) Western North Pacific monsoon, (c) East Asian monsoon, and (d) Australian monsoon regions.

5.1.3 Asymmetries with respect to El Niño and La Niña events

In the previous section, we evaluated the overall skill of four different MME methods in terms of seasonal mean temperature and precipitation over the globe and regional monsoon domain. In this section we investigated their performance during extreme weather events in certain years, (i.e., cases of strong El Niño and La Niña). However, we found that, in a comparison between SCM and SPM, the two weighted MME methods using multiple regression did not provide any improvement skills in the extreme years considered. Therefore, in this section, we only present the multi-model forecast skills, as obtained from SCM and SPM. The atmospheric responses to a developing and decaying ENSO event are seen to be nearly out of phase, especially for precipitation in the Asian-Australian monsoon (Wang *et al.* 2001; Wang *et al.* 2009). In order to investigate the ability of the multi-model predictions in presenting those features, composite maps of precipitation and 500 hPa geopotential height anomalies, normalized by their standard deviation, are shown in Fig. 10. The composites are calculated by using two El Niño (91/92, 97/98), three La Niña (88/89, 98/99, 99/00) and two normal (89/90, 93/94) boreal winters in observation, and a one-month lead, three-month mean prediction obtained from both SCM and SPM.

In observations, asymmetric patterns are clearly shown, with strong anomalies of precipitation and geopotential height in El Niño and La Niña, especially in the tropics where dryness (wetness) is shown over the Maritime continent and wetness (dryness) is shown over the tropical Pacific for El Niño (Na Nina) years. For both El Niño and La Niña events, the predicted normalized anomalies of precipitation and geopotential height, using the simple averaged MME, agree quite well with the corresponding observations; with high levels of the anomaly pattern correlation of more than 0.55 for precipitation and more than 0.70 for geopotential height (not shown) in both events, respectively. Significant errors are found over the Eurasian continent in both precipitation and geopotential height. However, the predicted anomalies of precipitation tend to be overestimated over most regions and a more symmetric pattern is shown for the anomaly between El Niño and La Niña. However, the predicted anomalies in both variables in normal years are weaker than in the observed years, and the value of the anomaly pattern correlation is modest, at 0.14, suggesting that without ENSO forcing the MME does not exhibit a particularly useful skill. Comparing the MME predictions, SPM shows an increased skill in both El Niño and La Niña events, with a higher pattern correlation (0.67) than SCM over the region. However, for normal years, the SCM prediction shows a slightly better skill than the SPM prediction. The pattern correlations of precipitation in both events over the Asian-Australian monsoon regions (IM, WNPM, EAM, and AM) are also investigated in Table 3 for a regional assessment. This clearly demonstrates that the corrected multi-model prediction is more effective in simulating the monsoon rainfall, even in strong El Niño and La Niña events, especially where the SCM prediction is relatively low (e.g., IM and EAM regions). For normal years, the simple averaged MME prediction performs better than the corrected one over all specific monsoon regions.

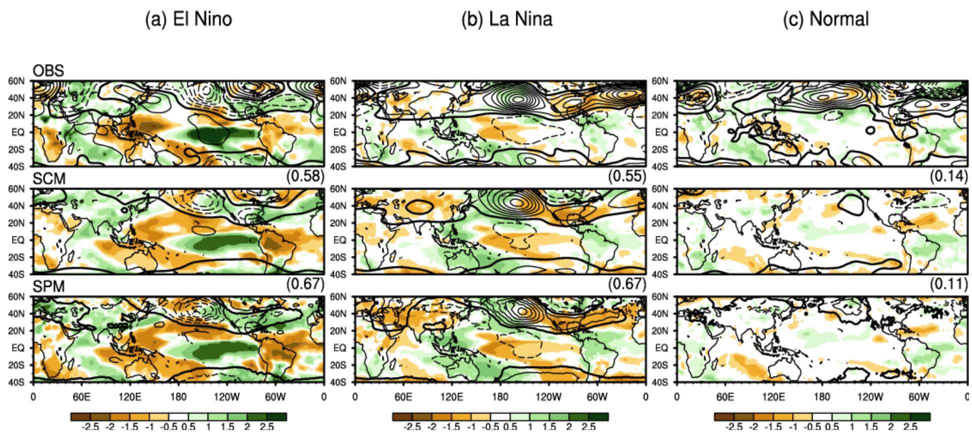

Figure 10 Precipitation (shaded) and 500hPa geopotential height (contour) anomalies composited for (a) El Niño (91/92, 97/98), (b) La Niña (88/89, 98/99, 99/00) and (c) normal (89/90, 93/94) boreal winters from observation (OBS) and multi-model predictions obtained by the SCM and SPM, respectively. All anomalies were normalized by their own standard deviations. The solid and dashed contours indicate the positive and negative anomalies of 500hPa geopotential height. Anomaly pattern correlations of precipitation are also displayed in the plot.

Table 3 Anomaly pattern correlations between the observed and predicted precipitation anomalies by the SCM and SPM for El Niño, La Niña, and normal DJF years over the monsoon regions

	El Niño		La Niña		Normal	
	SCM	SPM	SCM	SPM	SCM	SPM
Indian Monsoon	-0.18	0.13	0.29	0.70	0.05	-0.08
WNP Monsoon	0.91	0.96	0.89	0.91	0.47	0.16
East Asian Monsoon	0.79	0.85	0.50	0.90	-0.41	-0.57
Australian Monsoon	0.76	0.82	0.84	0.91	0.38	-0.04

5.2 Real-time forecast

To investigate the ability of empirically weighted and calibrated MME prediction systems in presenting seasonal mean temperature and precipitation for real-time forecasts, which is a very important issue from an operational perspective, we have evaluated the MME predictions for 12-running 3-month means during the period 2008–2011. Note that the APCC has started monthly 3-month mean forecasts since November 2007, although this time period is not sufficient to make comprehensive conclusions. Figures 11 and 12 show the skill of temperature and precipitation obtained from different MME methods for the independent real-time forecasts (2008–2011) in terms of anomaly pattern correlation over the large-scale regions; globe, northern hemisphere, tropics, and southern hemisphere. The averaged skill over the regions and their range of the highest and lowest skill scores for the whole time-series of forecasts (12 months × 4 years = 48 cases) obtained from the four MME prediction systems are also displayed.

As demonstrated in Figs. 4 and 5, it has also been shown that the general pattern of the MME skill is strongly related to the ENSO variability, in real-time forecasts (for the period of 2008JFM - 2011/12DJF). Correlations between the absolute value of the Niño 3.4 index and the averaged skill of four MME methods are 0.50 for temperature and 0.57 for precipitation, respectively. Moreover, similarly to the results from retrospective forecasts, weighted MME forecasts that are obtained by weighting the single-model prediction using multiple regression does not perform as well as either the simple averaged or the calibrated MME forecasts, across all the predictions for the whole period. There are few cases in which the skill of two weighted MMEs are better than that of SPM some cases but, as demonstrated in both retrospective and real-time forecasts, they are generally worse than the SCM and SPM skills considering all aspects of the predictions, also indicating the year-to-year skill variations for MRG and SSE suggested that the multiple regression-based schemes are unstable and exhibit larger range of skill than others. From Figs. 11 and 12, the positive impact of the model correction using the step-wise pattern projection method on seasonal mean temperature and precipitation can be summarized. Firstly, the SPM predictions perform consistently better than the others in real-time forecast periods, across all aspects of the variables, seasons, and large-scale regions, although the level of skill improvement with respect to the reference forecast is marginal (as compare with the retrospective forecast). Secondly, the year-to-year forecast skill of the SPM is generally more stable, indicating a narrow range of variations in anomaly pattern correlation.

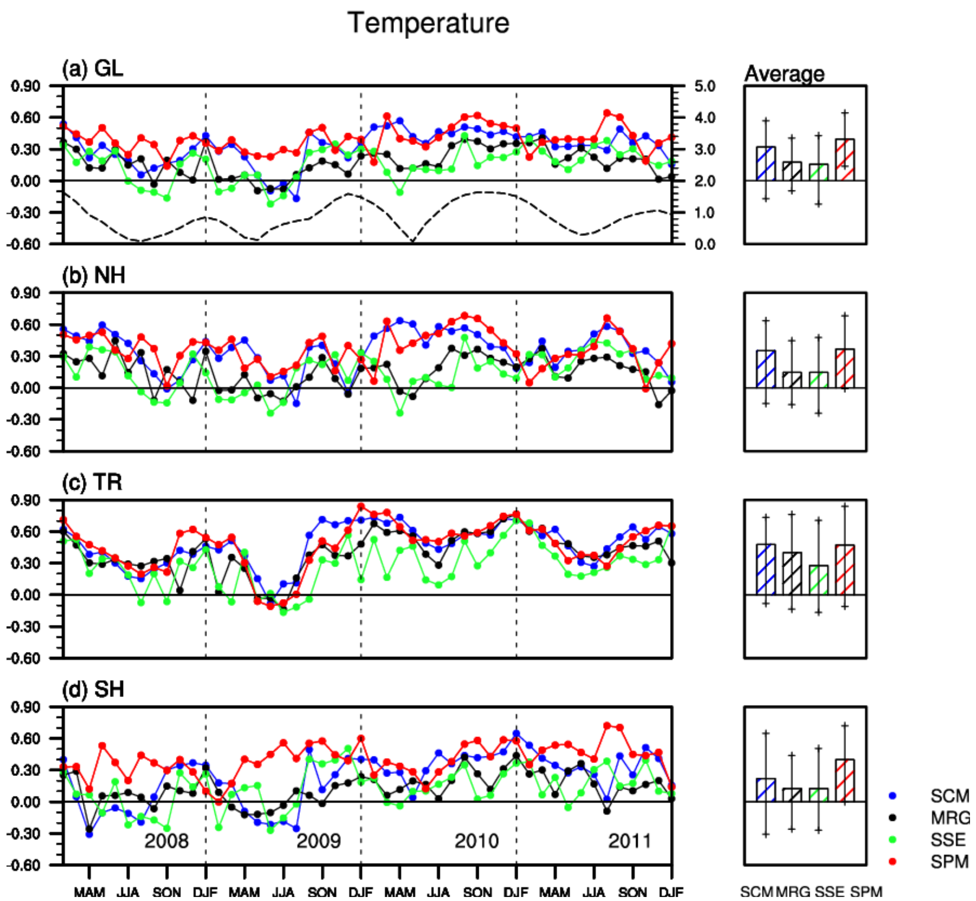

Figure 11 Time series of anomaly pattern correlation for 1-month lead seasonal mean temperature obtained from four MME methods (SCM, MRG, SSE, SPM) for the real-time period 2008–2011 over the globe, northern hemisphere, tropics, and southern hemisphere. The averaged scores and the range of scores for the whole series of forecasts are also shown. The black line indicates the amplitude of Nino 3.4 Index.

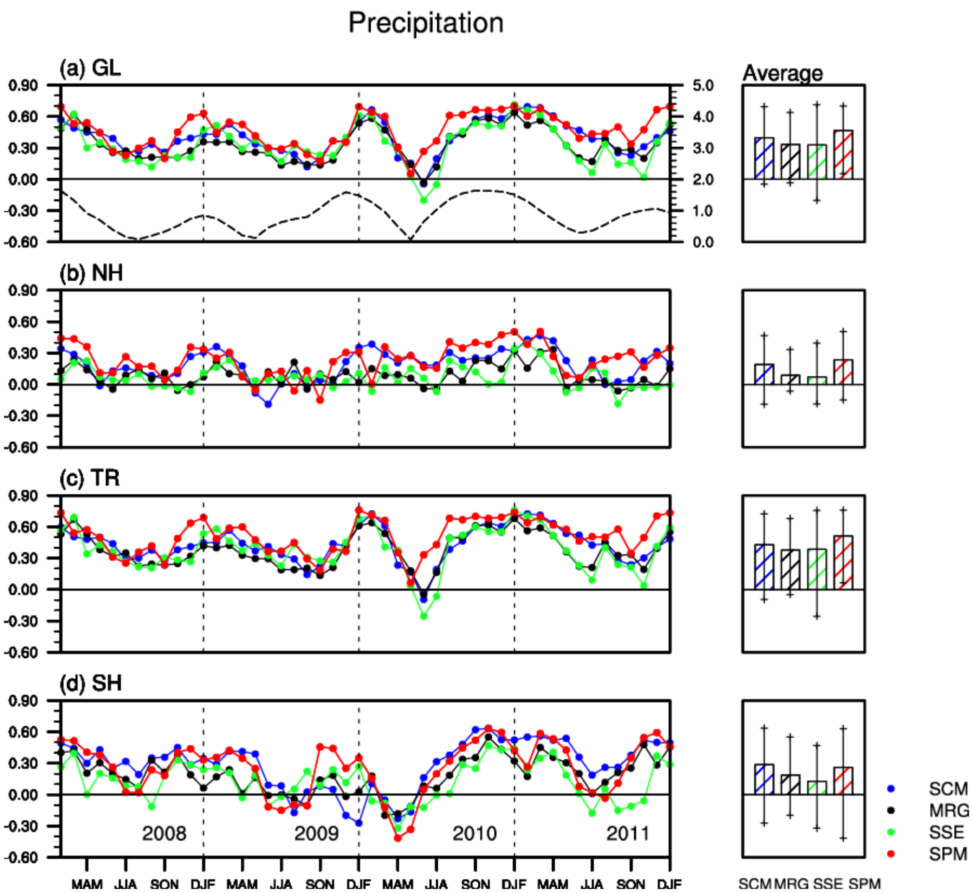

Figure 12 Same as Fig. 11, except for precipitation.

6. A BASIC FOR RATIONALITY OF MULTI-MODEL ENSEMBLE

Results demonstrated in the study and experience of the applications of the MME technology shows that the simple averaged multi-model prediction usually outperforms the schemes based on a regression method (e.g., Kharin and Zwiers 2002; Kryjov *et al.* 2006; DelSole and Shukla 2009; Wang *et al.* 2010). This may seem surprising, since regression, (and particularly, multiple regression) provides optimal weights of the predictors in the sense of minimization of the mean squared error and it is, therefore, an extremely popular and powerful statistical tool in empirical environmental research. When natural environmental variables are considered, multiple regression provides the best fit of the predictand within the training period, and even beyond the training period (in forecasting) when the overfitting problem is avoided. To understand the possible causes of the failure and success of the different MME schemes, we first investigated the sources of rationality of MME technology and evaluated to which degree the different MME methods fitted the rationality.

6.1 Simple averaged method

Let's consider ensemble mean forecast of a single-model X. It could be decomposed as:

$$X = kY + err\_m + err\_n \tag{1}$$

with k being the regression coefficient of the model ensemble mean forecast X on the observation Y. err_m is a model error, which is the same for all the ensemble members err_n is a standard error of mean, which depends on the ensemble spread and size and does not depends on the model error.

For greater transparency, Eq. (1) can be rewritten as:

$$X = kY + err \tag{2}$$

with the term err being the errors, which are a combination of the model error and the random error; $err = err\_m + err\_n$ (with err_n converging to zero with infinite increase of the ensemble size). The main advantage of this decomposition is that the two right-hand-side terms are not correlated with each other, and therefore the variance of X s equal to the sum of the variances of kY and the term err:

$$\sigma_x^2 = \sigma_{kY}^2 + \sigma_{err}^2 \tag{3}$$

The signal to error ratio (R) can therefore be expressed as follows:

$$R = \sigma_{kY}^2 / \sigma_{err}^2 \tag{4}$$

The main purpose of the multi-model approach is to increase this ratio.

In the above decompositions, k is a regression coefficient and, theoretically, there is no restriction on its value. However, when an environmental model is considered, it is obvious that this coefficient should be positive, as this is the essence of the term “model”. If a model commonly reproduced the environment with an inverse sign, it would be either be considered to be nonsense, or it could be attributed to the model errors being so large that they cancel out the signal and then, occasionally, the correlation could appear to be non-positive. An idea of the simple averaged MME with equal weighting (SCM) is based on Eq. (2), and there is an assumption that the regression coefficient k is positive for all the single-models involved in the MME prediction. Under such constraints however, in the MME forecasts, the signal fraction of variance increases, because values of kY are absolutely correlated in all the models; whereas the error fraction of variance decreases because the values of err are not absolutely correlated. It is easy to show that, in general, the signal to error ratio R in the MME is larger than in the individual model forecasts, and that it increases with the number of participating models. Certainly, it also depends on the individual model forecast variances and the individual model regression coefficient k. In this study, we will, however, consider them comparable for all the models.

To show that R increases with an increase in number of participating models, we must demonstrate that σ^2_{kr} increases faster in the sum of the model forecasts than σ^2_{err}. The variance of the sum of n series, $z_i(i=1,2,...n)$, is equal to the sum of the elements of the covariance matrix:

$$\sigma^2 \Sigma_{zi} = \sum \sigma^2_{zi} + 2\sum_{j>i} r_{ij}\sigma_i\sigma_j \tag{5}$$

where, r_{ij} is a correlation coefficient between the series z_i and z_j. The correlation coefficients between the signal terms of all the models are equal to one by definition, because the difference in signal terms between all the models is the constant positive factor k_i (i-index of the model). Therefore, in Eq. (5) the signal terms turns out as:

$$\sigma^2_{\sum k_i Y} = (\Sigma \sigma_{k_i Y})^2. \tag{6}$$

Meanwhile, the error terms are not absolutely dependent for all the models and, in general, correlation coefficients between the error terms differ from one. If the error terms are independent for all the models and non-correlated, Eq. (6) for them becomes:

$$\sigma^2_{\Sigma err_i} = \Sigma \sigma^2_{err_i}. \tag{7}$$

The square of the sum is larger than the sum of the squares by definition, so for the MME with independent errors the ratio R is larger than for the individual models, and it increases with the number of the participating models. If the model errors are correlated, but not absolutely, the values of the variance of the sum of the error terms increase faster than described by Eq. (7), but slower than described by Eq. (6).

It should be noted that it can be only considered under the above constraints. It is particularly important that the regression coefficients are positive and comparable for all the models; and the model variances are comparable for all the models. If all the above considerations are included, we can expect the value of R to increase with an increase in the number of the models. However, experience shows that when aggregated over large regions, an unlimited increase in the number of participating models does not lead to an unlimited improvement of the multi-model performance (e.g., Kharin and Zwiers 2002; Kug *et al.* 2008b). Instead, after the number of the participating models has reached a certain critical number, a further increase in the number of models does not contribute to the skill improvement; instead it converges to some value or even degrades. For example, Kharin and Zwiers (2002) demonstrated that the performance of the multi-model prediction becomes saturated using six or seven models, and even decreases with a further increase in the number of participating models.

The main cause of this saturation in real situations is that all the models combined in the MME do not match the above constraints for all the grid-points covered by the verification assessments. An application of simple linear regression-based methods is intent to avoid somehow those constraints. In particular, if the regression coefficients for all the models are positive and the statistical linear relationships between the model forecasts and the corresponding observations are stable for all the models during the training period, the application of the regressions to all the model forecasts can increase the weights of the models with a larger fraction of kY, which certainly, is reasonable. However, even in this the most obvious case, the degree to which the linear relationships described by the coefficient k will remain stable beyond the training period and will be applicable to the forecast, is not known. Furthermore, as practice shows, there are many grid-points for which the coefficient k is negative for all the participating models, especially in the extratropics, (see, for example, Fig. 13 that shows the regression coefficients of mean precipitation for JJA for each single model during the period 1983–2003). It should be noted that the paradigm beyond both described approaches, (either the simple averaged MME or the simple linear regression-based MME, with all the regression coefficients being positive), implies an enhancement of the signal (kY), which is absolutely correlated with observations in all the models. Note that this approach does not imply any correction of the individual model errors.

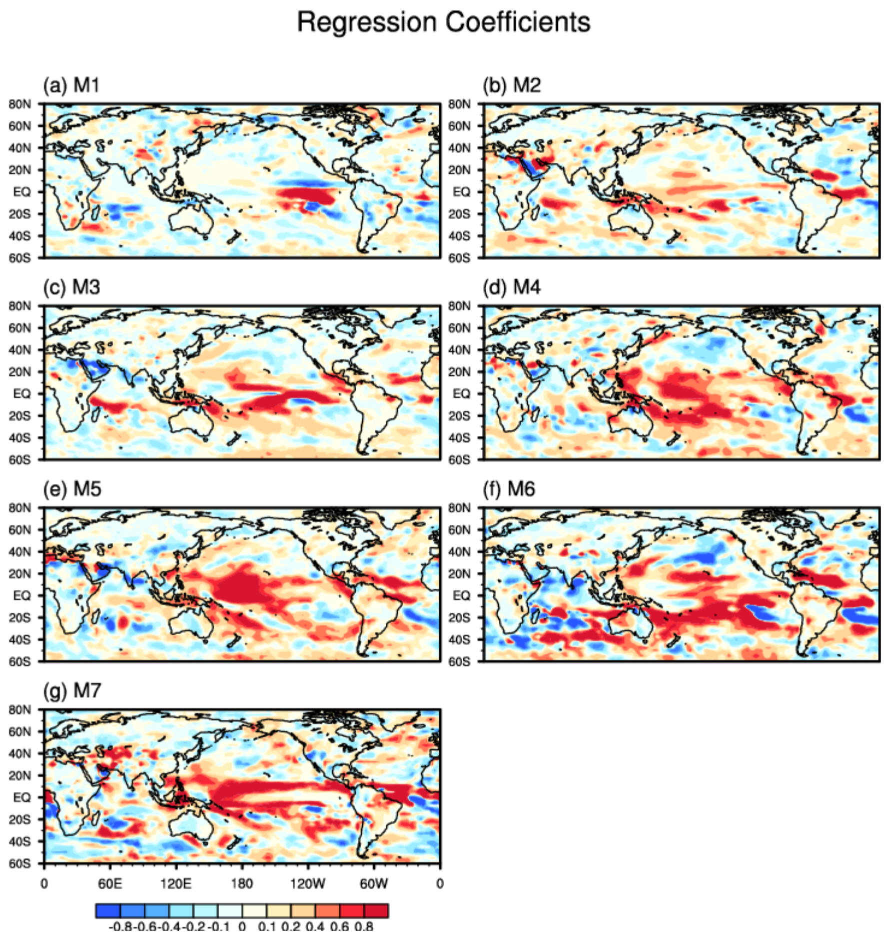

Figure 13 Weightings (regression coefficients) of summer mean precipitation for each single model obtained from regression analysis.

6.2 Point-wise multiple regression method

Multiple regression approaches, (e.g., MRG, SSE), differ in essence from the simple averaged and simple linear regression-weighted MME described above, where all the regression coefficients are positive. Note that in the simple linear regression-based MME, as discussed above, the regression coefficients are estimated separately for each model and the forecasts are then combined with the weights equal to the corresponding regression coefficients. However, the regression coefficients used in multiple regression are simultaneously estimated while accounting for other models. Instead of an enhancement of the signal (which is common in all the model forecasts) and a suppression of the errors (which is not common), multiple regression suggests a combination of all the signals and errors from the model forecasts which are to any degree partially corrected with observations, regardless of the sign, with the signal or error common to all the models being taken from a model with a large fraction of signals and errors.

To explain this, we shall consider partial coefficients which are the basis of multiple regression coefficients, and we shall consider the simplest case; a multiple regression which elates a predictand to two predictors, i.e., observations (Y) to the Model 1 forecasts (X_1) and Model 2 forecasts (X_2) A partial correlation coefficien $\rho_{YX_1X_2}$, which underlies the first regression coefficient, shows the linear relationship between observations to the Model 1 forecasts, with the influence from the Model 2 forecasts being removed:

$$\rho_{YX_1,X_2} = \frac{r_{YX_1} - r_{YX_2} r_{X_1X_2}}{\sqrt{1-r_{YX_2}^2}\sqrt{1-r_{X_1X_2}^2}} \tag{8}$$

The strength of the linear relationship between the first predictor and the observations is weakened by the degree to which it correlates with the second predictor and how the latter correlates with observations. The equation for the second multiple regression coefficient is similar. It means that from each predictor the multiple regression extracts only the information which is additional to other predictors, with that information which is common to all the predictors being extracted from only one predictor and not taken from the others. This approach is, therefore, opposite to the simple averaged MME and simple linear regression-based weighted MME, with all the regression coefficients being positive. Although the application of the multiple regression is appropriate for the assessments of linear relationships between the natural environmental variables, when each additional predictor adds some variability not noticed by the leading predictor, it appears to be completely unreasonable to link the natural environmental variables (predictands) with the model variables, because there is a common addition of erroneous information from which multiple regression extracts “useful signals.” However, the signal extracted from the errors cannot be stable and cannot add value to the forecast.

In addition, when dealing with the multiple regression, the overfitting problem needs to be considered. we have to consider the overfitting problem. During the training period, with the addition of each new involved predictor, the multiple regression curve fits closer to the predictand curve. However, overfitting during the training period results in a failure of the forecasts. Estimation of too many regression coefficients leads to overfitting that causes a degradation in skill (Davis 1976; Michaelsen 1987), and the rule of thumb recommends not more than one predictor for each 10 to 20 years of the training period (e.g., Chetyrkin and Kalikhman 1982). It means that for a 25–30 year training period, which is ordinary nowadays, there must not be more than 2–3 model predictors. However, nowadays it is not unusual for the number of participating models in the MME (as predictors in multiple regression) to be between 8 and 10. Tests to avoid the overfitting problem performed by DelSole (2007) show that even with the use of a successful solution, by, for instance, the use of the (multiple) ridge regression, the forecast is not improved. The author explained the need for a more precise tuning of the ridge regression for each grid-point. However, it is likely that the main cause of the failure of the multiple regression applications in the combining of the MME, resides in the mismatch between the paradigms of the multiple regression and the MME, as discussed above, rather than the overfitting problem.

6.3 Spatial shift correcting method

Another approach to the MME combination resides in the spatial correction of the models forecasts and the MME combination of the same. The current seasonal climate models still have systematic and random errors that degrade seasonal climate predictions, as reported in many studies (e.g., Feddersen *et al.* 1999; Kang *et al.* 2004; Wang *et al.* 2008; Jin *et al.* 2008; Kug *et al.* 2008b; Alessandri *et al.* 2010; Lee *et al.* 2010; Lee and Wang 2012). Errors in the anomaly component are related to the incorrect performance of climate models in simulating the anomalies, and the model anomalies tend to be misplaced when contrasted with the observed dataset

(Kang and Shukla 2006). The spatial shifts of the simulations can be corrected by statistical correction (or downscaling) methods, in particular, those based on a linear correlation between the model and the observed patterns; a so-called pattern projection technique (e.g., Kang *et al.* 2004; Feddersen *et al.* 2005; Kang and Shukla 2006) or by using a technique suggested by Min *et al.* (2011).

As demonstrated earlier, the calibrated MME prediction system, SPM, is one of the spatial shift correcting methods that uses a simple averaged MME with equal weightings from corrected single-model predictions using the step-wise pattern projection method. The reason why the calibrated MME predictions consistently show a better and more stable skill than others, for both the retrospective and real-time forecasts, is because the current dynamical climate models are capable of capturing large-scale patterns related to local variability, although they have difficulty in predicting local variability correctly at each grid point. Therefore, it has been demonstrated that the SPM, by correcting the (spatially shifted) model errors using a statistical correction technique, consistently shows a better and more stable skill than others for both the retrospective and the real-time forecasts, when considering all aspects of the predictions.

7. SUMMARY

Since 2007, the APCC has provided one-month lead three-month mean MME predictions for global temperature and precipitation and disseminated these to APEC member economies. Four different MME prediction systems are currently operationally implemented for deterministic (or ensemble-mean) seasonal forecasts; simple averaged MME with equal weightings (SCM), multiple regression-based weighted MME (MRG), multiple regression-based MME with EOF-filtered datasets (SSE), and calibrated MME with corrected single-model predictions using the step-wise pattern projection method (SPM). One of the purposes of the present study was to provide a comprehensive documentation of the seasonal forecasts, issued by the APCC operational MME prediction systems, with a large set of predictions. In order to investigate the ability of the weighted and calibrated MME prediction systems, we assessed their overall performance using 12 running three-month mean temperature and precipitation forecasts with respect to the reference forecast (SCM), for the retrospective (1983–2003) and real-time (2008–2011) periods.

The results indicate that multiple regression-based weighted MMEs (MRG and SSE) generally do not perform as well as the simple averaged MME and the calibrated MME over most latitudinal zones, for all variables and seasons. The identity of the multiple regression-based weighted MMEs varies depending on which aspect of the forecast is considered; they are, therefore, considered to be unstable. Although in certain individual cases the MRGs and SSEs perform better than the SCMs and SPMs, it is evident that the overall conclusion, considering all aspects of the predictions, is still valid. The results also clearly demonstrate that the model correction using the step-wise pattern projection method (SPM) has a positive effect in improving the multi-model predictions, with a more extensive coverage of high levels of skill, compared with the other models. In particular, the SPM' ability to predict temperature is relatively greater over the tropics and the southern hemisphere, while for precipitation the model correction contributes to the increase in forecast skill in most latitudinal zones across all seasons. Similar results, (from the large-scale assessment), can be found over the Asian-Australian monsoon regions for a regionalized assessment, in which the APCC is highly interested. Therefore, the calibrated MME prediction by model correction is the most effective way in reducing errors and improving forecasts for all the regional monsoon domains (i.e., Indian Monsoon, East Asian Monsoon, Western North Pacific Monsoon, East Asian Monsoon and Australian Monsoon), in terms of anomaly pattern correlation – RMSE diagram. Moreover, the predicted normalized anomalies obtained by calibrated MME prediction system capture the observed asymmetric patterns between a strong El Niño and La Niña quite well, with a higher pattern correlation than using any other reference system.

To better understand the possible causes of the failure and success of the different MME schemes, we investigated the sources of rationality of the MME technology and evaluated to which degree different MME methods fit the rationality. If all the models contain a useful signal, the simple averaged MME provides the best results. With reference to the grid-points for which all the models contain a useful signal, (with the power of the useful signals essentially varying between the models), it could be possible to improve the skill of the MME by using simple linear regression-based weights, given that the regression relationships are stable. However, in real situations, there are too many grid-points in which the regression coefficients are negative and unstable, because robust weights are difficult to calculate given the relatively short samples available in training the regression models (e.g., Kharin and Zwiers 2002; Peng *et al.* 2002; Doblas-Reyes *et al.* 2005; Kug *et al.* 2008b). In addition, multiple regression-based weighted MMEs do not provide a greater improvement in skills than simple averaged MME, due to, not only the signal problem, but also the existence of errors common to all the models and to the problem of overfitting. Therefore, the use of the simple averaged MME is a practical way of utilizing the multi-model approach in an operational environmental. However, the skill of the model can be improved by correcting the spatial shift of the single-model errors using a statistical correction method and then combining with equal weights to issue a multi-model prediction. Thus improved, the MME indicates a consistently better and stable skill than the simple averaged and the multiple regression-based MME schemes.

REFERENCES

Alessandri A, Borrelli A, Masina S, Pietro PD, Carril A, *et al.* (2010) The INGV-CMCC seasonal prediction system: Improved ocean initial conditions. Mon Weather Rev 138:2930-2952

Alexander MA, Blade I, Newman M, Lanzante JR, Lau NC, Scott JD (2002) The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans. J Clim 15:2205-2231

Back S K, Ryu JH, Ryoo SB (2002) Analysis of the CO2 doubling experiment using METRI AGCM part I: The characteristics of regional and seasonal climate response. J Korean Meteor Soc 38:465-477

Barnston AG, Li S, Mason SJ, Dewitt DG, Goddard L, Gong X (201) Verification of the first 11 years of IRI's seasonal climate forecast. J Appl Meter Clim 49:493-520

Boer GJ (2005) An evolving seasonal forecasting system using Bayes'theorem. Atmos-Ocean 43:129-143

Chetyrkin EM, Kalikhman IL (1982) Probability and Statistics. Finansyi Statistika, Moscow (in Russian)

Davis RE (1976) Predictability of sea level pressure anomalies over the North Pacific Ocean. J Phys Ocean 6:249-266

DelSole T (2007) A Bayesian framework for multimodel regression. J. Clim 20:2810-2826

DelSole T, Shukla J (2009) Artificial skill due to predictor screening. J Climate 22:331-345. doi: 10.1175/2008JCLI2414.1

Ding Y, Ni Y, Zhang X, Li M, Dong W, Zhao ZC, Li Z, Shen W (2000) Introduction to the short-term climate prediction model system. China Meteorological Press, Beijing, China (in Chinese)

Doblas-Reyes FJ, Deque M, Piedelievre JP (2000) Multi-model spread and probabilistic seasonal forecasts in PROVOST. Q J R Meteorol Soc 126:2069-2088. doi: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

Feddersen H, Andersen U (2005) A method for statistical downscaling of seasonal ensemble predictions. Tellus 57A: 398–408

Feddersen H, Navarra A, Wrad MN (1999) Reduction of model systematic error by statistical correction for dynamical seasonal prediction. J Clim 12: 1974-1989

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

Janowiak JE, Xie P (1999) CAMS_OPI: a global satellite-raingauge merged product for real-time precipitation monitoring applications. J Clim 12: 3335-3342

Jia X, Lin H, Lee JY, Wang B (2012) Season-dependent forecast skill of the dominant atmospheric circulation patterns over the Pacific North-American region. J Clim in press. Doi: 10.1175/JCLI-D-11-00522.1

Jin EK, Kinter JL, Wang B, Park CK, Kang IS et al (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere models. Clim Dyn 31:647-664

Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL (2002) NCEP-DOE AMIP-II reanalysis (R-2). Bull Am Meteorol Soc 83: 1631-1643

Kang H, Park CK, Hameed SN, Ashok K (2009) Statistical downscaling of precipitation in Korea using multimodel output variables as predictors. Mon Wea Rev 137: 1928–1938. doi:10.1175/2008MWR2706.1

Kang IS, Ho CH, Lim YK, Lau KM (1999) Principal modes of climatological, seasonal, and intraseasonal variations of the Asian summer monsoon. Mon Weather Rev 127:322-340

Kang IS, Lee JY, Park CK (2004) Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. J Clim 17:834-844

Kang IS, Shukla J (2006) Dynamical seasonal prediction and predictability of the monsoon. Wang B (ed) The Asian Monsoon, Chap 15, Springer, pp 585-612

Kharin VV, Zwiers FW (2002) Climate predictions with multimodel ensembles. J Clim 15:793-799

Kharin VV, Zwiers FW (2003) Improved seasonal probability forecast. J Clim 16: 1684-1701

Krishnamurti TN, Kishtawal CM, Shin DW, Williford CE (2000) Multimodel ensemble forecasts for weather seasonal climate. J Clim 13:4196-4216. doi:10.1175 / 1520-0442(2000)013 〈 4196:MEFFWA 〉 2.0CO;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

Kryjov VN *et al.* (2006) Assessment of the climate forecasts produced by individual models and MME methods. APCC Technical Report 2006, APEC Climate Center, Korea

Kug JS, Lee JY, Kang IS (2008a) Systematic error correction of dynamical seasonal prediction of sea surface temperature using a stepwise pattern projection method. Mon Weather Rev 136:3501-3512

Kug JS, Lee JY, Kang IS, Wang B, Park CK (2008b) Optimal multi-model ensemble method in seasonal climate prediction. Asia-Pacific J Atmos Sci 44:259-267

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

Lau NC, Nath MJ (1996) The role of atmospheric bridge in linking tropical Pacific ENSO events to extratropical SST anomalies. J Clim 9:2036-2057

Lee DY, Ashok K, Ahn JB (2011a) Toward enhancement of prediction skills of MME seasonal prediction: A climate filter concept. J Geophys Res. doi:10.1029/2010JD014610

Lee JY (2003) Assessment of potential seasonal predictability with a multi-model dynamical-statistical ensemble system. Ph.D. thesis, Seoul National University, Seoul, Korea, 164pp

Lee JY, Wang B (2012) Seasonal climate prediction and predictability of atmospheric circulation. In Leonard M. Druyan (Ed.) Climate Models, InTech Open Access Books, 19-42

Lee JY, Wang B, Kang IS, Shukla J et al (2010) How are seasonal prediction skills related to models' performance on mean state and annual cycle? Clim Dyn 35: 267–283. doi:10.1007 / s00382-010-0857-4

Lee JY, Wang B, Kang IS, Shukla J, Kumar A, Kug JS, *et al.* (2010) How are seasonal prediction skills related to models' performance on mean state and annual cycle? Clim Dyn 35:267-283. doi:10.1007/s00382-010-0857-4

Liou C S, Chen J. H, Terng CT, Wang FJ, Fong CT, Rosmond TE, Kuo HC, Shiao CH, Cheng MD (1997) The second generation global forecast system at the central weather bureau in Taiwan. Wea. Forecasting 3:654-663

McFarlane NA, Boer GJ, Blanchet JP, Lazare M (1992) The Canadian climate centre second generation

circulation model and its equilibrium climate. J Clim 5:1013-1044
Michaelsen J (1987) Cross-validation in statistical climate forecast model. J Clim Apply Meteor 26:1589-1600
Min YM, Kryjov VN, Oh JH (2011) Probabilistic interpretation of regression-based downscaled seasonal ensemble predictions with the estimation of uncertainty. J Geophys Res. doi:10.1029 / 2010JD015284
Min YM, Kryjov VN, Park CK (2009) A probabilistic multimodel ensemble approach to seasonal prediction. Wea Forecasting 24: 812-828
Min YM, Lee JY, Kug JS, Wang B, Kryjov VN et al (2012) Improvement of the APCC multi-model probabilistic seasonal prediction by model calibration and combination. Clim Dyn (revision)
Palmer TN, Alessandri A, Andersen U et al (2004) Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). Bull Am Meteorol Soc 85:853-872
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
Park H, Park BK, Rha DK, Cho JY (2002) An improvement of global model in 2001, KMA/NWPD Technical Report, 2002-1 (in Korean)
Park HS, Ahn JB (2004) Development of a new CGCM and ENSO Hindcast Experiment using the CGCM(1). Asia-Pacific J Atmos Sci 40:135-146
Peng P, Kumar A, Barnston AG, Goddard L (2000) Simulation skills of the SST-forced global climate variability of the NCEP-MRF9 and Scripps-MPI ECHAM3 models. J Clim 13:3657-3679
Peng P, Kumar A, vad den Dool H, Barnston AG (2002) An analysis of multi-model ensemble predictions for seasonal climate anomalies. J Geophys Res, 107. doi:10.1029/2002JD002712
Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. J Clim 15:1609-1625
Ritchie H (1991) Application of the semi-Lagrangian method to a multilevel spectral primitive-equations model. Q J R Meteorol Soc 117:91-106
Saha SS, Nadiga S, Thiaw C et al (2006) The NCEP Climate Forecast System. J Clim 19:3483-3517
Scinocca JF, Mcfarlane NA, Lazare M, Li J, Plummer D (2008) The CCCma third generation AGCM and its extension into the middle atmospheres. Atmos Chem and Phys 8:7055-7074
Shukla J et al (2000) Dynamical seasonal prediction. Bull Am Meteorol Soc 81:2493-2606. doi:10.1175/1520-0477 (2000) 081 〈2593:DSP〉 2.3.CO;2
Sohn SJ, Min YM, Lee JY, Tam CY et al (2012) Assessment of the long-lead probabilistic prediction for the Asian summer monsoon precipitation (1983-2011) based on the APCC multimodel system and a statistical model. J Geophys Res, 117, D04102
Takaya Y, Yasuda T, Ose T, Nakaegawa T (2010) Predictability of the mean location of typhoon formation in a seasonal prediction experiment with a coupled general circulation model. J Meteor Soc Japan88:799–812
Wang B, Kang IS, Shukla J, Lee JY, *et al.* (2010) Improvement of APCC seasonal prediction and assessment of characteristics and forecast of 2009/2010 climate anomalies. Final Report of APCC International Research Project 2010, APEC Climate Center, Korea

Wang B, Lee JY, Kang IS, Shukla J et al (2008) How accurately do coupled climate models predict the leading modes of Asian-Australian monsoon interannual variability? Clim Dyn 30:605-619. doi:10.1007/s00382-007-0310-5

Wang B, Lee JY, Kang IS, Shukla J, Park CK et al (2009) Advance and prospectus of seasonal prediction: Assessment of APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980-2004). Clim Dyn 33:93–117

Wang B, Lin Ho (2002) Rainy season of the Asian-Pacific summer monsoon. J Clim 15:386-398

Wang B, Wu R, Lau KM (2001) Interannual variability of Asian summer monsoon: contrast between the Indian and western North Pacific-East Asian monsoon. J Clim 14:4073-4090

Wilks DS (1995) Statistical Methods in the Atmospheric Sciences: An Introduction. Academic Press, 467 pp

WMO (2002) Standardised Verification System (SVS) for Long-Range Forecasts (LRF). New attachment II-9 to the manual on the GDPS. Vol 1 WMO-No 485, 24 pp

Yoo JH, Kang IS (2005) Theoretical examination of a multi-model composite for seasonal prediction. Geophys Res Lett 32: L18707. doi:10.1029/2005GL023513

Yun WT, Stefanova L, Krishnamurti TN (2003) Improvement of the multimodel superensemble technique for seasonal forecasts. J Clim 16: 3834-3840

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

APCC TECHNICAL REPORT 2012-01

- Application of Bayesian Model Averaging on Multi-Model Ensemble Seasonal Prediction
- Decadal Change of Variability and Predictability of Two Types of ENSO
- Assessment of Relationship between EL Nino and Indian Summer Monsoon Rainfall
- Long-lead MME Extreme Drought Prediction
- Assessment of APCC Multi-Model Ensemble Predictions

APEC Climate Center

12, Centum 7-ro, Haeundae-gu, Busan 612-020,
Republic of Korea
Tel: +82-51-745-3900 Fax: +82-51-745-3949
www.apcc21.org

비매품
94500
9 788997 333363
ISBN 978-89-97333-36-3
ISBN 978-89-97333-35-6 (세트)