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 2011, 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 largest Multi-Model Ensemble (MME) system for seasonal prediction through its international science network and has provided value-added products to various stakeholders. Recently, APCC has expanded its mandate to include enhancement of the capacity of APEC member economies information to respond effectively to climate change and variability through better application of climate.

To achieve its research and social objectives, in 2011, APCC made efforts to research improvements in its climate prediction performance from various angles and towards better understanding of climate variability and the reproducibility of the climate models for the relevant application of climate information to society. The following technical report provides more information about our research outcomes from 2011.

APCC will continue to improve the quality and accuracy of 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 results for the benefit of society and academia. We also 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

Improvement of the APCC Probabilistic Multi-Model Ensemble Prediction by Model Calibration and Combination

■■ Dr. Young-Mi Min

APCC

TECHNICAL REPORT 2011-02

Improvement of the APCC Probabilistic Multi-Model Ensemble Prediction by Model Calibration and Combination

Dr. Young-Mi Min

ABSTRACT

Efforts have been devoted to improve the probabilistic multi-model prediction (PMMP) system that is operationally employed in the Asia-Pacific Economic Cooperation (APEC) Climate Center (APCC). The novelty of the proposed system lies in (i) the use of an upgraded multi-variable version of a stepwise pattern projection method (SPM) to calibrate single model predictions and obtain a more reliable forecast probability, and (ii) the combination of skillful models based on the selection of such models from among calibrated single-model predictions to formulate multi-model probabilistic prediction. The former first corrects errors in the predicted anomalies and then inflates the variance of the corrected prediction to match that of the corresponding observed variance in the individual models. The latter produces a tercile-based categorical PMMP based on the combination of skillful models from among all possible candidates after the calibration.

A comprehensive assessment of the benefits of the calibration (using the SPM and variance inflation) and combination (based on skill-based model selection) in the new PMMP system was first carried out for 23-year retrospective forecasts (1981-2003) of temperature and precipitation in a double cross-validation mode. The results indicate that both the calibration and combination significantly contribute to improving the PMMP skill for most regions of the globe. The resolution of the forecasts is improved by model correction and combination, while reliability is mainly increased by inflation. As a result, the calibrated PMMP based on model combination has significantly higher skill than the current version of (uncalibrated) the operational PMMP system for both variables over the globe. It is further shown that the new PMMP system also improves forecast skill during the real-time forecast period of 2008-2010 relative to the performance of the current operational system.

1. INTRODUCTION

The Asia-Pacific Economic Cooperation (APEC) Climate Center (APCC) has made an effort, since its inception in 2005, to facilitate the share of high-cost climate data and information to minimize the economic and human losses due to natural disasters. The APCC has also served to enhance capacity building by meeting the growing societal and economic interests in the monitoring and prediction of seasonal climate variability. Since 2007, the APCC has produced one-month lead three-month mean forecasts every month with deterministic (based on ensemble mean) and probabilistic (based on ensemble spread) interpretations from a well-validated multi-model ensemble

(MME) seasonal prediction system, and disseminated it to APEC member economies. 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.

The MME technique was designed near the turn of this century to quantify forecast uncertainties due to model formulation (Krishnamurti et al. 1999, 2000; Doblas-Reyes et al. 2000; Palmer et al. 2000; Shukla et al. 2000) and has since been considered to be an effective method to improving weather and climate forecasts. In general, the MME-based prediction is superior to that made by any single-model component for both two-tier (Krishnamurti et al. 1999, 2000; Palmer et al. 2000; Shukla et al. 2000; Barnston et al. 2003; Wang et al. 2004) and one-tier systems (Hagedorn et al. 2005; Doblas-Reyes et al. 2005; Yun et al. 2005; Wang et al. 2008, 2009; Kug et al. 2008a; Lee et al. 2010, 2011b, c).

The APCC has devoted considerable effort to developing a multi-model prediction system for producing improved and well-validated seasonal forecasts. Currently, four methods for constructing a multi-model ensemble from single-model ensembles are operationally used for deterministically interpreted (hereafter, deterministic) seasonal forecasts. The first method is a simple averaged MME where the contribution of each single-model is equally weighted (i.e., a simple composite method; SCM). The second method is a calibrated MME which is obtained from the adjusted (or corrected) single-model predictions based on a stepwise pattern projection method (SPM; Kug et al. 2008c). Another approach involves using empirically weighted MMEs with coefficients computed using multiple linear regression (MRG; Krishnamurti et al. 2000; Yun et al. 2003), and MRG can also be completed with an empirical orthogonal function (EOF)-filtered dataset to minimize the residual error variance (i.e., a synthetic superensemble method; SSE; Yun et al. 2005). Wang et al. (2010) demonstrated that, in general, the SPM has better forecast skill than the other three methods for both the retrospective and real-time forecast periods.

Climate forecasts are associated with uncertainty due to the chaotic nature of the climate system, and the level of forecast uncertainty can be quantified by a

probability distribution function (PDF) (Zwiers 1996; Kharin and Zwiers 2001). Probabilistic forecasts not only provide more useful information on the uncertainty by displaying chaotic nature of the climate system but are also of greater value to end users for decision-making as compared to deterministic ones (Kryzsztofowicz 1983; Cusack and Arribas 2009; Alessandri et al 2011). As a result, various probabilistic multi-model prediction (PMMP) systems have been employed at several operational centers that routinely provide MME seasonal forecasts (e.g., APCC, ECMWF[1], IRI[2], MSC[3]). However, the majority of studies on calibrated MME methods are focused on deterministic interpretation (Yun et al. 2003, 2005; Kug et al. 2008c; Ke et al. 2009; Lee et al. 2011a) despite the fact that the superiority of the multi-model concept is more clearly evident in a probabilistic framework (Doblas-Reyes et al. 2000; Palmer et al. 2004). Some efforts in this direction have been practically done at a few operational centers, e.g., IRI (Barnston et al. 2003) and CPTEC[4] (Coelho et al. 2006). However, although the calibrated probabilistic forecasts are more related to the user application to match the needs of regional climate risk management or only issuing for a particular variable/region of their interest with only a few exceptions, most of operationally implemented PMMP systems generally provide non-calibrated global seasonal forecasts.

The operational probabilistic forecasts at the APCC are also based on an uncalibrated MME with model weights that are inversely proportional to the random errors in the forecast probability associated with the standard error of the ensemble mean, i.e., model weights that are proportional to the square root of model ensemble size (Min et al. 2009). The APCC issues seasonal forecasts in the form of tercile-based categorical probabilities, that is, the probabilities of below-normal (BN), near-normal (NN), and above-normal (AN) categories. The tercile-based categorical probabilities from an ensemble of forecasts, including the tercile information based on the hindcasts, are estimated by a parametric estimator derived from a fitted Gaussian distribution (e.g., Kharin and Zwiers 2003a; Boer 2005; Min et al. 2009; Min et al. 2011). The

1) The European Centre for Median-range Weather Forecasts, UK
2) The International Research Institute for climate society, USA
3) Meteorological Service of Canada, Canada
4) Centro de Previsão de Tempo e Estudos Climáticos, Brazil

pattern of probabilistic forecast produced by the APCC operational PMMP system is quantitatively similar to that of the forecast anomalies produced by the SCM, although the probabilistic forecasts take into account the model uncertainty.

Motivated by the superiority of the SPM over other deterministic MME methods (e.g., Kang and Shukla 2006; Kug et al. 2008c), this study aims at improving an operational version of the uncalibrated PMMP system by calibrating single-model predictions using the SPM and variance inflation for probabilistic interpretation. The inflation is applied to enhance the estimation of forecast uncertainty, specifically by adjusting (or rescaling) the variance of the SPM-based corrected predictions to match the corresponding observed variance for accurate estimation of forecast uncertainty. Both techniques are used here as calibration strategies separately for each single-model predictions. Note that the version of the SPM used here differs that of the previous studies of Kug et al. (2008c). Our method includes an upgraded multi-variable SPM with an increased number of potential predictors, whereas the previous method is based on a mono-variable version. Moreover, it is the first effort to apply a SPM-style method to probabilistic predictions, while the previous studies were focused only on deterministic predictions. In order to construct a multi-model prediction from the corrected single-model predictions, a skill-based model selection approach is used in the study.

To investigate the benefits of the proposed methods for model calibration and combination, a set of experiments is carried out to verify forecast skill for temperature at 850hPa (hereafter, temperature) and precipitation for boreal summer seasons (June-July-August, JJA) for the period of 1981-2003. Based on the proposed methods, we develop a calibrated PMMP system and compare it with the current operational version of the PMMP system. The paper is organized as follows. Section 2 describes the dynamical climate models used in the study and their retrospective/real-time forecast datasets and corresponding observations, and also provides a brief description of the experiments and verification metrics. The methods employed to calibrate and combine the single-model predictions are introduced in Section 3. In Section 4 we present the results from the single-model calibration and combination, illustrating the advantages of the proposed methods. Section 5 describes the overall performance

of the calibrated PMMP system in comparison with that of the uncalibrated operational system. A brief summary and discussion of the results follows in Section 6.

2. DATA AND EVALUATION OF FORECAST SKILLS

2.1 Model and observed data

The models examined in this study come from 10 dynamical climate prediction systems from APEC member economies that currently participate in the APCC operational one-month lead seasonal mean MME predictions every month. Table 1 lists the acronyms of the institutes and models used here. Table 2 presents a brief summary of model specifications for seven two-tier (CWB, GCPS, GDAPS, GEM, AGCM2, SEF, and NIMR) and three one-tier (NCEP, PNU, and POAMA) systems, respectively. Note that three models from MSC, the CCCma second generation atmospheric general circulation model (AGCM2; McFarlane et al. 1992), a reduced-resolution version of the medium-range weather forecast global spectral model (SEF; Ritchie 1991) and the global environmental multiscale model (GEM; Cote et al. 1998) developed at the RPN, are independent of each other[5]. The models show a large range of model resolutions and ensemble sizes and their retrospective forecast datasets match the requirements of the Seasonal Prediction Model Intercomparison Project-2/Historical Forecast Project (SMIP2/HFP). All models have generated ensemble retrospective forecasts for the common period of 1981-2003 and real-time forecasts for the common period of 2008-2010. In the present study, we focus on one-month lead seasonal mean forecasts of JJA temperature and precipitation. The model dataset is interpolated to a common grid resolution of 2.5olon x 2.5olat, which is similar to the observed data grid.

The data used for verification of the retrospective and real-time forecasts of temperature were obtained from the NCEP-Department of Energy (DOE) reanalysis 2 data (Kanamitsu et al. 2002). The Climate Anomaly Monitoring System (CAMS)

5) Information available at http://www.weatheroffice.gc.ca/saisons/howto_seasonal_0-3_e.html

and Outgoing longwave radiation Precipitation Index (OPI) data (Janowiak and Xie 1999), which produces real-time monthly analyses of global precipitation merged with satellite rainfall estimates from the OPI with ground based rain gauge observations from the CAMS, was used as verification data for precipitation.

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

Acronym	Full Names
AGCM2	Second Generation Atmospheric General Circulation Model
BMRC	Bureau of Meteorology Research Center
CCCma	Canadian Centre for Climate Modeling and Analysis
CWB	Central Weather Bureau
GCPS	Global Climate Prediction System
GDAPS	Global Data Assimilation and Prediction System
GEM	Global Environmental multiscale Model
KMA	Korea Meteorological Administration
MSC	Meteorological Service of Canada
NCEP	National Centers for Environmental Prediction
NIMR	National Institute of Meteorological Research
PNU	Pusan NationalUniversity
POAMA	Predictive Ocean-Atmosphere Model for Australia
RPN	Recherche en Prévision Numérique du temps
SEF	Spectral aux éléments finis Model

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

Model Acronym	Institution (Country)	Resolution	Ensemble Size	SST Specification (Hindcast/Forecast)	Reference
CWB	CWB (Chinese Taipei)	T42 L18	10	Predicted SST/Predicted SST	Liou et al. (1997)
GCPS	SNU (Korea)	T63 L21	12	Predicted SST/Predicted SST	Kang et al. (2004)
GDAPS	KMA (Korea)	T106 L21	20	Predicted SST/Predicted SST	Park et al. (2002)
GEM	MSC (Canada)	2°x2° L50	10	Persistent ERA40[6]-SST/ Persistent CMC[7] SST	Cote et al. (1998)
AGCM2	MSC (Canada)	T32 L10	10	Persistent ERA40-SST/ Persistent CMC SST	McFarlane et al. (1992)
SEF	MSC (Canada)	T95 L27	10	Persistent ERA40-SST/ Persistent CMC SST	Ritchie (1991)
NIMR	NIMR (Korea)	5°x4° L17	10	Persistent OISST[8]/ Persistent OISST	Back et al. (2002)
NCEP	NCEP (USA)	T62 L64	15	Predicted SST/Predicted SST	Saha et al. (2006)
PNU	PNU (Korea)	T42 L18	5	Predicted SST/Predicted SST	Park et al. (2004)
POAMA	BMRC (Australia)	T47 L17	10	Predicted SST/Predicted SST	Zhong et al. (2005)

2.2 Forecast skill measures

The metrics used to measure the prediction skills of the deterministic forecasts include the temporal correlation coefficient (TCC) and the anomaly pattern correlation coefficient (PCC). The Student's t-statistic (*t*-test) was used to assess the statistical significance of the estimated TCC at the 5% level. Following the recommendations of the World Meteorological Organization Standardized Verification System for Long-Range Forecast (SVS-LRF; WMO 2002), the probabilistic forecasts were assessed by using relative operating characteristics (ROC; Mason 1982; Wilks 1995; Richardson 2000; Zhu et al. 2002) and reliability diagrams including the unconditional frequency distribution of the forecasts (Murphy 1973; Wilks 1995; Atger 2003; Jolliffe and Stephenson 2003). The ROC curves measure the ability of the forecasts to detect the occurrence and non-occurrence of a seasonal climate event, thus measuring

6) ECMWF 40 year Re-Analysis
7) Canadian Meteorological Centre
8) National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation SST

resolution, and the area under of the ROC curve (ROC score) is a commonly used statistic representing the skill of the probabilistic forecasts. The reliability diagram displays the relative frequency of an observed event against the forecast probability of the event for each bin. That is, the reliability diagram provides information on how closely the forecast probabilities of an event correspond to their observed frequencies, thus indicating the reliability of the forecasts.

Following the recommendations of the WMO SVS-LRF, the ROC curves/scores and the reliability diagrams were spatially aggregated over large regions to estimate large-scale verification statistics in order to evaluate the overall skill of the forecast system and ultimately assess its improvement (WMO 2002). The globe (0-360°E, 60°S-60°N) and two sub-regions; the tropics (0-360°E, 20°S-20°N) and East Asia Monsoon region (EAM; 110°-140°E, 20°-45°N), are used in the study. In addition to the aggregated verification, a grid-point verification in terms of the ROC score was performed for a regionalized assessment of the skill of the forecast system. The statistical significance of the estimated ROC scores was assessed using the Monte Carlo (MC) simulation by scrambling the forecast dataset with replacement in the time domain 500 times (Stephenson and Doblas-Reyes 2000; Min et al. 2009). For each of 500 MC trails, we estimated all the verification scores for the grid points and then assessed the significance as the probability of the randomly obtained verification scores exceeding those obtained from the original succession of forecasts. We considered the ROC score statistically significant at the 5% level. A similar procedure was applied to the significance tests of the difference between the two ROC scores.

To assess the skill of the probabilistic forecasts with respect to the climatological forecast in each event (i.e., AN, or NN, or BN category), we also used the Brier Score (BS) and Brier Skill Score (BSS) including its decomposition in reliability (Brel) and resolution terms (Bres; Wilks 1995). The Brier score (BS) is the most commonly used measure of accuracy of a probabilistic forecast and the definition of the BS is as follows:

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

where, n is the number of forecasts, f_i the forecast probability of occurrence for the ith forecast, and o_i is the ith observed probability, which is defined to be 1 if the event occurs and 0 otherwise. The BS can be decomposed as three terms (Wilks 1995):

$$BS = \bar{o}(1-\bar{o}) + \frac{1}{n}\sum_{k=1}^{m} n_k (f_k - \bar{o}_k)^2 - \frac{1}{n}\sum_{k=1}^{m} n_k (o_k - \bar{o}_k)^2 = BS_{unc} + BS_{rel} - BS_{res}$$

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. As indicated, the three terms are known as uncertainty, reliability, and resolution. The BSS is defined with respect to a reference forecast (e.g., a climatological forecast):

$$BSS = 1 - \frac{BS}{BS_{ref}} = \frac{BS_{res} - BS_{rel}}{BS_{unc}}$$

The BSS is unity for a perfect forecast and zero or negative for forecasts that are unskillful relative to the reference forecast.

2.3 Experimental design

In this study, we used the model anomalies to remove the climatological model biases (i.e., systematic errors in the climatology or mean bias of the forecast). For each individual model, the anomalies are estimated as departures from their climatology over the training period in a one-year out cross-validation way (Wilks 1995). This implies that a new climatology is calculated at each cross-validation step, with the target year being withheld. That is, the anomalies are estimated using climatology from the 22-year long training periods for the retrospective forecasts of 1981-2003 and using the whole 23-year climatology for the real-time forecasts of 2008-2010. The same procedure of estimation of anomalies is applied to the observed dataset.

Prior to applying the proposed calibration and combination methods to the current

operational version of the uncalibrated PMMP system, a range of experiments for the retrospective forecast of 1981-2003 were carried out to assess the benefits of the different single-model and multi-model calibration methods, as well as of the single-model combination for MME. The experiments will be described in Section 3. The performance of the newly designed PMMP system has been evaluated for its prediction skill for JJA mean temperature and precipitation, presented in the form of tercile-based categorical probabilities, and compared with the performance of the uncalibrated PMMP system (Min et al. 2009). In this study, only two types of events have been considered for verification; predictions above the upper and below the lower tercile (i.e., AN and BN categories[9]), for both the 23-year retrospective and 3-year real-time forecasts, which are the common periods for all of the models listed in Table 2. Note that, for the 23-year retrospective forecast (1981-2003), skill was calculated using the one-year out cross-validation method based on the 22-year training period with the target year withheld. As for the independent real-time forecasts (2008-2010), we used the whole 23-year hindcast dataset as the training period for each real-time forecast.

3. METHODS FOR THE PMMP IMPROVEMENT

3.1 Model correction and combination

The current seasonal climate models, even the state-of-the-art coupled models, still have systematic and random errors that degrade seasonal climate prediction, as many studies have described (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 incorrect performance of climate model 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

9) Results for the NN category are not shown in this study because its probability is relatively insensitive to signal perturbations. As a result, the probability of the NN category is close to the climatological probability, thus the outcome for the NN event cannot be predicted with high confidence, as found in many previous studies (e.g., van den Dool and Toth 1991; Kharin and Zwiers 2003a).

of the simulations can be corrected by statistical correction methods based on the linear correlation between the model and observed patterns, so-called pattern projection technique (e.g., Kang et al. 2004; Feddersen et al. 2005; Kang and Shukla 2006).

Because the SPM has been shown to be superior to other methods in correcting the errors in model anomalies in a deterministic framework (e.g., Kang and Shukla 2006; Kug et al. 2007; Kug et al. 2008b, c), we adopted an improved SPM-based correction method to improve the PMMP system. The SPM is a kind of point-wise regression model based on pattern projection and it is based on the large-scale patterns of the predicted variables by individual models (i.e., predictors) correlated with a local (or grid) observed variable (i.e., predictand). Kang and Shukla (2006) and Kug et al. (2008b and c) demonstrated that the pattern projection method is more effective in improving seasonal prediction than other commonly used statistical correction methods based on maximum covariance analysis (sometimes referred to as singular value decomposition).

The SPM is designed 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. The model equation is as follows:

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

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

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

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

where x and t represent spatial and temporal grid points, respectively. $X(t)$ indicates a time series projected by the covariance pattern between predictand ($Y(t)$)

and predictor field ($\psi(x,t)$) in a certain domain (D). The covariance pattern indicates a pattern of the model prediction which is related to the observed predictand. The parameter, α, 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 between the predictand and the two-dimensional predictors at each grid point during a training period are calculated. The correlation coefficients at each grid point are first categorized into several groups depending on their absolute values; 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, and 0.1. If the number of grid points in the first group (i.e., $\geq|0.8|$) is larger than a criteria, the grid points are used for the predictor domain. If not, the next group in descending order (i.e., $\geq|0.7|$) is included in the predictor domain together with the grid points in the first group. As a result, the grid points that are relatively well-correlated during the training period are selected as the predictor domain. In this regard, the SPM differs from the traditional statistical models with one fixed geographical domain, and the predictor domain can be automatically selected from grid points in several regions. Based on the selected predictor domain, a statistically corrected prediction is produced by pattern projection, according to the aforementioned equation. In our case, the criterion is 100 grid points same as Kug et al. (2008b). The criterion for groupings and the number of grid points is somewhat arbitrary. However, Kug et al. (2008b) showed that it is not sensitive to the choice of the criterion. Note that the above statistical correction procedures are carried out in one-year out cross-validation mode and applied separately for each single-model ensemble-mean prediction. For more details on the SPM procedures, refer to Kug et al. (2008b and c).

To improve the predictive skill of the SPM, we increase the number of potential predictors. In the previous version of the SPM implemented in the APCC operations for deterministic MME forecast (Kug et al. 2008c), the predicted temperature (precipitation) field is used as a single predictor to correct temperature (precipitation) anomaly at each grid point. On the other hand, this study uses the multi-variable version of the previous one and we have restricted the pool of potential predictors for following two reasons. First, predictors should be available from 10 of the

dynamical climate models used in the study (Table 2). Second, the predictors should be well simulated by the models. Based on the results of the correlation analysis between the various potential predictors and corresponding observations, two more large-scale variables; 500-hPa geopotential height (Z500) and sea level pressure (SLP), are used in the study as additional predictors (see Fig. A1 in Appendix A). These variables are commonly used as potential predictors for statistical correction or downscaling studies (e.g., von Storch and Zwiers 1999; Kang et al. 2002; Wetterhall et al. 2005; Chu et al. 2008; Jia et al. 2011; Min et al. 2011). For each model, the SPM-based statistical correction method is carried out separately for each predictor, and the final prediction is obtained as a simple average of the corrected (ensemble-mean) predictions from the three predictors. The corrected single-model and multi-model predictions using the mono- and multi-variable versions of the SPM for each grid point will be referred to in the following as experiments 'SPM-mono' and 'SPM-multi', respectively (Table 3).

Prior to creating the multi-model prediction from the corrected single-model predictions, a model selection method that considers model performance is applied for each grid point. In a double cross-validation mode (Feddersen et al. 1999; Min et al. 2011), a subset of models for multi-model combination at a target year is selected based on their past performance of the corrected predictions with the observed anomalies during the hindcast period, by withholding of the target year. For example, to select the models for the year 1981, we assess the skill of the predictions for the period 1982-2003 and the predictions for each year are obtained using the SPM based on the 21-year hindcast dataset as the training period for the statistical correction model. The model selection procedure is repeated for each target year and grid point. In our case, the skill threshold used is the TCC value of 0.3, which approximately corresponds to the 5% significance level in the one-tailed Student t-test. In this regard, the combination method has an advantage of preferentially excluding some models with hindcast skill below the threshold for a given variable and grid-point. Based on the corrected single-model predictions from the skill-based selected models, the simple averaged MME prediction (with equal weighting) is constructed for each grid point. This experiment is denoted hereafter as 'COM' (Table 3).

Table 3 Set of model calibration and combination experiments used in the study.

Experiment	Acronym	Description
None	RAW	Single (or multi)-model ensemble from raw model output
Correction	SPM-mono	Corrected single (or multi)-model ensemble using the mono-variable version of SPM
	SPM-multi	Corrected single (or multi)-model ensemble-mean using the multi-variable version of SPM
Combination	SCM	Simple multi-model ensemble (with equal weighting)
	COM	Simple multi-model ensemble obtained from the SPM-based corrected single-model predictions of the skill-based selected models
Inflation	INF	Inflated simple multi-model ensemble obtained from the SPM-based corrected single-model predictions of the skill-based selected models (i.e., inflated COM)

3.2 Variance inflation and probabilistic approach

Linear regression models, which are widely used for statistical correction, are based on least squares fit and almost inevitably lead to a loss of variability in the reconstructed or predicted field (Feddersen et al. 1999; von Storch 1999). This well-known problem with regression-based corrected model forecasts is commonly overcome by calibrating the corrected field using a simple inflation method (e.g., Klen et al. 1959; Karl et al. 1990; Huth 1999; Kang et al. 2004). In a probabilistic forecast, an accurate estimation of forecast probabilities is important for quantifying the forecast uncertainty (Kharin and Zwiers 2003a; Tippett et al. 2007; Min et al. 2011). Therefore, an inflation of the ensemble spread is required in order to obtain reliable probabilities (Hamill and Colucci, 1998; Doblas-Reyes et al. 2005). Recently, with increasing use of ensemble predictions, a scheme accounting for the ensemble spread has been suggested by Feddersen and Andersen (2005). Most recently, Min et al. (2011) introduced a new approach to estimate total forecast uncertainty originating from both regression and ensemble spread of model forecasts within the regression framework based on error analysis (Taylor 1982).

With the above as motivation, prior to probabilistically interpreting the corrected ensemble-mean predictions, we applied an additional calibration method to rescale (or adjust) the variance of the predictions to match the corresponding observed

variance for each single model. The standard inflation method used in the study is to multiply the adjusted values (i.e., regression-based corrected predictions at each grid point) by the ratio between the standard deviations (SDs) of the observations and those of the adjusted values. That is, the inflation modifies the predictions to have the same interannual variance as the observed dataset at every grid point (Doblas-Reyes et al. 2005). The inflated (or rescaled) prediction, $Y_{IF\_SPM(k,t)}$, in a grid point k at a certain year t is defined here as

$$Y_{IF\_SPM(k,t)} = Y_{SPM(k,t)} \times \frac{\sigma_{OBS(k)}}{\sigma_{SPM(k)}},$$

where $Y_{SPM(k,t)}$ is the SPM-based corrected ensemble-mean in a grid point k at a certain year t; $\sigma_{OBS(k)}$ and $\sigma_{SPM(k)}$ are the SDs of the observations and adjusted fields at a grid point k during the training period, respectively. It should be noted that the inflation factor is also cross-validated. The inflation of the single-model (ensemble-mean) prediction was conducted in the experiment denoted as 'INF' (Table 3).

Based on the calibrated multi-model predictions using the above described methods, the tercile categorical forecasts were considered. A parametric Gaussian fitting method was used to estimate probabilistic information from an ensemble of deterministic forecasts (e.g., Kharin and Zwiers 2003a; Boer 2005; Min et al. 2009, 2011). In this study, we used the calibrated single-model predictions (using the SPM and variance inflation) as members of an ensemble to take into account model uncertainty. The forecast probability of each category was estimated as a portion of the forecast PDF with respect to the observed climatological one. That is, the first of the PDF parameters, the mean, was constructed with the simple average of the corrected single-model predictions of the skill-based selected models. The second parameter, forecast uncertainty, was estimated as the model spread of the corrected predictions and then as the inflated single-model predictions of the selected models at each grid point. In this regard, there might be some grid points where no single-model predictions contribute to the MME due to their poor skills for the

hindcast period, even with statistical correction. In cases where no models were found to benefit the MME, a climatological distribution was taken to estimate forecast probability.

4. SENSITIVITY TEST FOR THE PROPOSED METHOD

4.1 Model correction and combination

In this section we present the performance of the different methods described in Section 3 based on retrospective forecasts for the period of 1981-2003. Figure 1 shows the averaged TCC over the globe of the single- and multi-model predictions obtained using the different versions of the SPM with respect to the raw model output (RAW). Here, the estimated scores are first calculated for each grid-point and then averaged over the globe. The SCM is the simple averaged MME with equal weighting from the ten single-model predictions (M1-M10). As expected, the SCM predictions generally perform better than any single-model predictions, as it has been reported (Doblas-Reyes et al. 2005; Hagedorn et al. 2005; Min et al. 2009; Wang et al. 2009). First of all, we examined the general performance of the corrected single-model and multi-model predictions using the previous version of the SPM (SPM-mono). The SPM-mono is capable of improving JJA mean temperature and precipitation predictions by correcting the errors in model anomalies for all the single models (only except for temperature prediction by M8) and their corresponding SCM, supporting conclusions of Kug et al. (2008c). Furthermore, the SPM-mono precipitation predictions for a few models are as skillful as or even more skillful than the RAW SCM prediction. On average, the positive impact of single-model calibration using the SPM-mono is relatively larger than that of multi-model calibration. Especially, a considerable improvement of forecast quality is shown by the models for which the RAW predictions have poor skills.

In order to test the benefits of the upgraded version of the SPM (SPM-multi), we compared its forecast skill with that of the SPM-mono. In general, the SPM-multi

skill is slightly higher than that of the SPM-mono for both temperature and precipitation. There are just few cases in which the single-model performance of the SPM-mono is better than or comparable to that of the SPM-multi. However, it should be noted that the main advantage of using the SPM-multi in the study is not a significantly large improvement in a certain case, but rather the consistently better performance of the SPM-multi when considering all aspects of the single-model and multi-model predictions for both temperature and precipitation over the globe.

To explore how model combination based on the skill-based selection method can affect the performance of the multi-model prediction, the COM results are also displayed with respect to the SCM results in Fig. 1. As described in Section 3, the COM is a method of constructing a multi-model prediction by selecting of a subset of models based on their hindcast skill of the corrected predictions prior to multi-model combination and then simply averaging them with equal weighting. By comparing the corrected SCM and COM predictions, it is evident that skill is increased when only the skill-based selected models are considered for multi-model combination. This result is valid for both versions of the SPM and illustrates the advantage of model combination. Moreover, it clearly demonstrates that the combined method of model correction, which employs the SPM-multi and model combination based on the skill-based selected models, is the most effective correction method for JJA mean temperature and precipitation over the globe. This result is confirmed by the estimated TCCs being statistically significant at the 5% level from the one-tailed Student t-test. In the MME experiment for temperature, on average, the degree of increased TCC due to model correction (from the comparison of the RAW and corrected SCM predictions) is comparable with that due to model combination (from the comparison of the corrected SCM and COM predictions). However, for precipitation, model combination contributes more to increased forecast skill than does model correction.

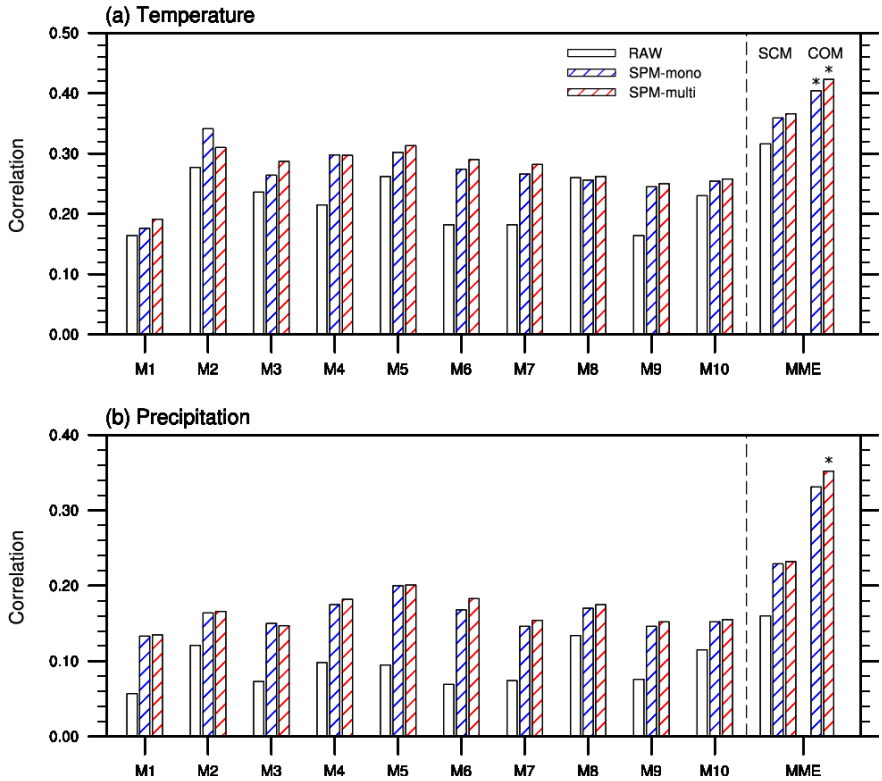

Figure 1 TCC for ten single-model ensembles (M1-M10) and multi-model ensemble (MME) predictions of JJA mean (a) temperature and (b) precipitation over the globe for the period of 1981-2003. The open bars represent the scores of the raw model output (RAW). The blue and red dashed bars correspond to the results obtained from the corrected (ensemble-mean) predictions using the mono- and multi-variable versions of the SPM (SPM-mono and SPM-multi). The SCM and COM are obtained from the simple averaged MME with equal weighting from all the ten single models and from the selected models based on their hindcast skill of the corrected predictions in double-cross validation mode, respectively. The asterisk represents the estimated TCC being statistically significant at the 5% level using the one-tailed Student's t-test.

Figure 2 describes the relative difference between the SCM TCC and COM TCC that is calculated as a difference between them divided by the SCM TCC for each SPM and variable separately. That is, a positive difference corresponds to the COM TCC being higher than the SCM TCC, and the degree to which the skill is higher is indicative of the efficiency of model combination in improving the corrected multi-model prediction. As shown in Fig. 1, the TCC of the corrected COM prediction is generally superior to that of the corrected SCM TCC, especially for precipitation.

It is worth noting that the impact of model combination on the improvement of the multi-model prediction is larger for the SPM-multi than for the SPM-mono, which is evidenced in Figure 2 by the differences "A" for temperature and "B" for precipitation. This implies that model correction using the SPM-multi and model combination work together to increase of forecast skill.

The largest positive impact of the COM can be found for the SPM-multi prediction for precipitation, the skill of which is up to 54% greater than that of the SPM-multi SCM prediction. As a result, the averaged TCC over the globe for precipitation is up to two times greater for the SPM-multi COM prediction than for the RAW SCM prediction (Fig. 1 (b)). The considerable skill enhancement may be attributable to the overall performance of the models in predicting precipitation generally being lower than for other climate variables (e.g., temperature, Z500, SLP), as reported in many studies (e.g., Peng et al. 2000; Derome et al. 2001; Wang et al. 2009; Lee et al. 2011b, c). Therefore, there are some limitations to correcting the spatial shifts of simulated precipitation by using the SPM-mono based on a single-predictor, precipitation. Moreover, because large-scale climate patterns play an important role in regulating local-scale precipitation, we use statistical methods to correct for model errors. Due to its increased number of potential predictors, relative to the SPM-mono, the SPM-multi gives a more skillful prediction in terms of the averaged TCC over the globe. Additionally, by preferentially excluding models with hindcast skills below a threshold, only skillful models can contribute to the MME in this study. As a result, the MME shows considerable improvement in skill scores, compared to previous methods, when both the SPM-multi and COM are included, particularly for precipitation.

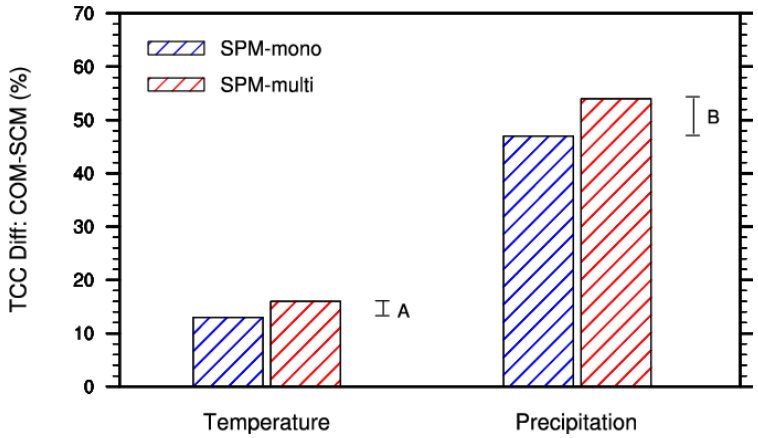

Figure 2 Relative difference between the SCM and COM TCC of JJA mean temperature and precipitation over the globe for the period of 1981-2003. The blue and red dashed bars correspond to the results obtained from the corrected multi-model predictions using the SPM-mono and SPM-multi. The relative difference is calculated by the difference between the SCM and COM TCC divided by the SCM TCC for each version of the SPM and variable separately. A and B indicate the increase in relative difference of the SPM-multi COM prediction with respect to the SPM-mono COM prediction for temperature and precipitation, respectively.

Predictability is usually higher over the tropical regions than in the extratropics, as shown in many studies (e.g., Kharin and Zwiers 2003a; Palmer et al. 2004; Min et al. 2009; Wang et al. 2009; Lee et al. 2011b, c). Therefore, an improvement of predictability for the extratropics may be more meaningful than for the tropics (Doblas-Reyes et al. 2005). Figure 3 shows the zonally averaged TCC of the SCM and COM predictions which have been corrected using the SPM-mono and SPM-multi, respectively. For comparison, the RAW SCM prediction is also displayed. Results indicate that both model correction and combination generally increase the skill for most of the latitudinal zones. For temperature, the positive impact of model correction using the SPMs (with respect to the RAW) is relatively larger for the tropics (20°S-20°N), while that of model combination (with respect to the SCM) is relatively larger for the extratropical regions in both hemispheres. On the other hand, both model correction and combination consistently improve the precipitation prediction for most of the latitudes, although the latter gives superior improvement for several regions (i.e., in the equatorial region (10°S-10°N), the area around 20°N and the high latitudes of both hemispheres (>50°N/S)). This also supports the result shown in Fig. 1(b) that the increased skill for precipitation prediction due to model

combination is relatively larger than that due to model correction in terms of the globally averaged TCC. Finally, it can also be shown that the SPM-multi has consistently better performance, particularly over the extratropical regions, than the SPM-mono, although the difference is quite marginal.

In summary, the combined method of model correction using the SPM-multi and combination is significantly superior to the separate methods for both variables over most of the latitudinal zones, as indicated by most of the TCCs estimated by the combined method being statistically significant at the 5% level from the one-tailed Student t-test across the globe. As compared with the reference multi-model predictions (i.e., the RAW SCM predictions), the major improvement due to the combined method is relatively larger over the tropical regions for temperature and over the subtropical and the extratropical regions for precipitation. Moreover, the calibrated multi-model prediction using the SPM-multi and COM combination shows not only consistently better performance across the globe, but also more stable forecast skill than the other methods, as indicated by the range of the highest and lowest TCCs being smaller across the considered latitudes (60°S-60°N).

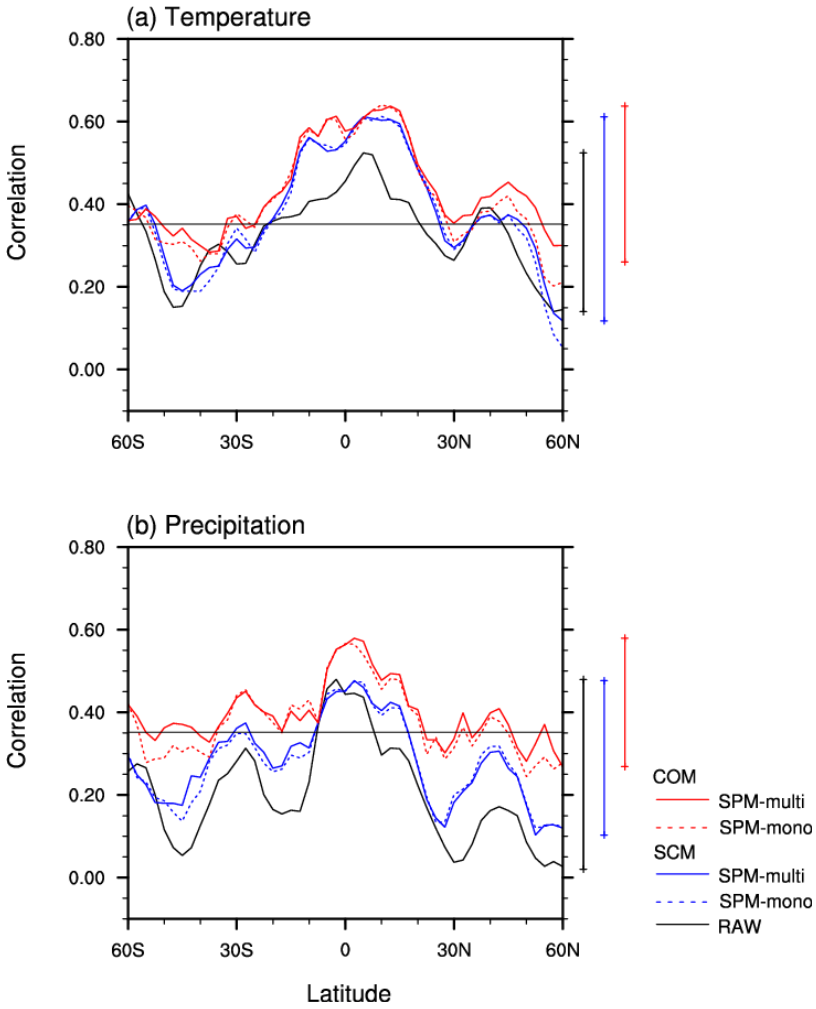

Figure 3 Zonal mean TCC of the corrected SCM and COM predictions using the SPM-mono and SPM-multi for JJA mean (a) temperature and (b) precipitation for the period 1981-2003. The solid (dashed) blue and red lines correspond to the scores obtained from the corrected SCM and COM prediction using the SPM-multi (SPM-mono). The solid black line represents the simple MME from the raw model output (RAW). The vertical black, blue, and red lines indicate the range of the highest and lowest TCCs across the globe, obtained from the RAW SCM, SPM-multi SCM, and SPM-multi COM predictions, respectively. The horizontal black lines correspond to the estimated TCC being statistically significant at the 5% level using the one-tailed Student's t-test.

4.2 Variance inflation

Previous experiments have shown the advantage of model correction using the SPM-multi and combination methods based on the skill-based model selection in the formulation of multi-model prediction. As described in Section 3, in order to compensate for the loss of variability in the regression models, the INF experiment has been applied in the cross-validation mode to the SPM-multi single-model predictions of the selected models (COM). In order to examine how well variance inflation can correct the underestimation of the predicted interannual variability (IAV), we compared the SDs of the predicted anomalies from the COM and the inflated COM (INF) with the observed SD (OBS) for the period of 1981-2003 (Fig. 4). In the observation, JJA mean temperature anomalies show a large variability in the high latitudes of the northern hemisphere and the eastern tropical Pacific along the west coast of South America. The predicted spatial pattern by the COM appears to be quite similar to that of the corresponding observations, with a considerably high PCC of 0.94, but the COM significantly underestimates the observed amplitude. After applying the INF to the COM, the method's ability to simulate the amplitude of the observed IAV is remarkably improved, as evidenced by its PCC value of 0.99. In contrast to temperature, a large variability of observed precipitation is only confined to the Asian monsoon regions, the western tropical Pacific, and the intertropical convergence zone along the equatorial Pacific. As with temperature, the underestimation of the COM's IAV is also featured by precipitation and it can also be mostly corrected by the INF.

To investigate how the variance inflation of the corrected single-model predictions can affect the probabilistic forecast skill, we have assessed it in terms of BSS including its decomposition in reliability (Brel) and resolution terms (Bres). Figure 5 shows the zonally averaged BSS, Brel and Bres of the probabilistic multi-model predictions in the AN categories obtained from the COM and INF. The figure indicates that the inflation effectively increases reliability but not resolution, as also shown by Kharin and Zwiers (2003a, b) and Doblas-Reyes et al. (2005). A general improvement of BSS is found for the INF and this result is valid for most of the latitudes and variables. It can also be seen that the positive impact of the INF for precipitation is quite

marginal for most of the latitudes, while the increase in BSS for temperature, which is mainly due to the increase in reliability, is greatest for the high latitudes of the northern hemisphere (>30°N) and middle latitudes of the southern hemisphere (20°S-40°S). This result is in agreement with the result shown in Fig. 4, suggesting that the INF is effective over the regions where the corrected prediction tends to underestimate the observed IAV. To summarize, the inflation of the model forecasts can contribute to the accurate estimation of the forecast uncertainty and then to the increase in forecast reliability that leads to the improvement of the probabilistic forecast skill. Note that these results are also valid for the BN categorical forecasts (not shown).

To explore the benefits of the model correction using the SPM-multi and combination in a probabilistic forecast, the reference skill of the AN categorical forecast from the raw multi-model predictions (RAW) is also displayed in Fig. 5. While the improvement of the reliability is mainly due to the inflation (from a comparison of the INF and COM in Brel in Fig. 5), the increase in resolution can be considered a consequence of model correction and combination (from a comparison of INF/COM and RAW in Bres in Fig. 5). A similar result was found by Stefanova and Krishnamurti (2002), who showed that the resolution terms are the most important contributors to the improvement of regression-based multi-model predictions over simple regression-based models in a probabilistic forecast. Given the improvement of both the reliability (due to variance inflation) and the resolution (due to model correction and combination), the model correction and combination can improve BSS relative to the reference case since it increases both resolution and the reliability.

Based on our sensitivity tests of the different experiments (Table 3), we suggest that the combined method of model calibration (using the SPM-multi and variance inflation) and combination (based on the skill-based selected model) is the most effective way to improve multi-model probabilistic prediction. This finding provides a basis for the consideration of the use of the proposed methods for improving the operational version of the uncalibrated PMMP system, which is discussed in the following Section.

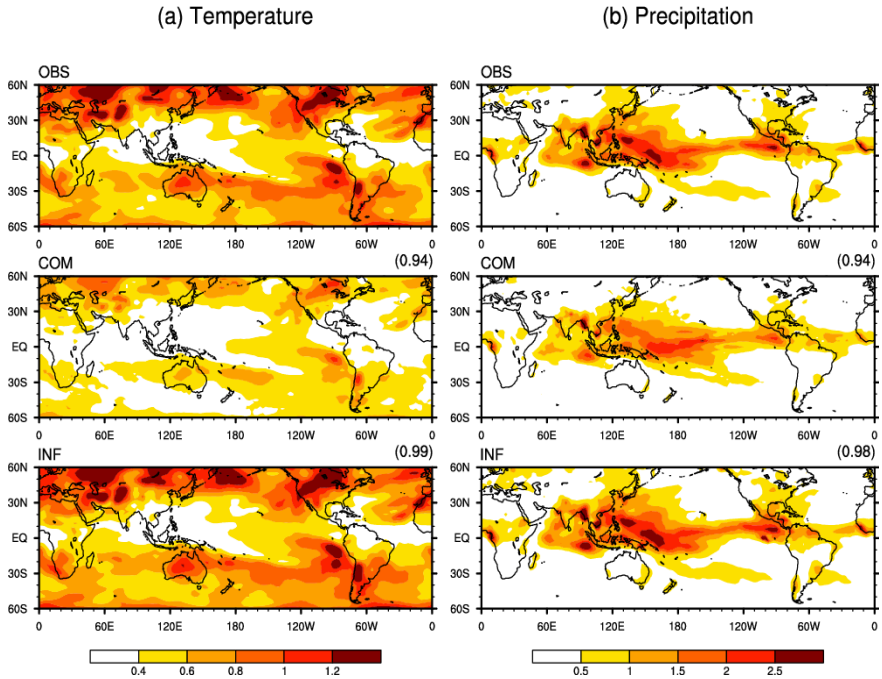

Figure 4 Standard deviation of interannual variability of JJA mean (a) temperature and (b) precipitation anomalies for the period 1981-2003 from the observation (OBS), the SPM-multi multi-model predictions of the skill-based selected models (COM), and the inflated COM (INF). The parenthesized value indicates the anomaly pattern correlation coefficient with respect to the observed one.

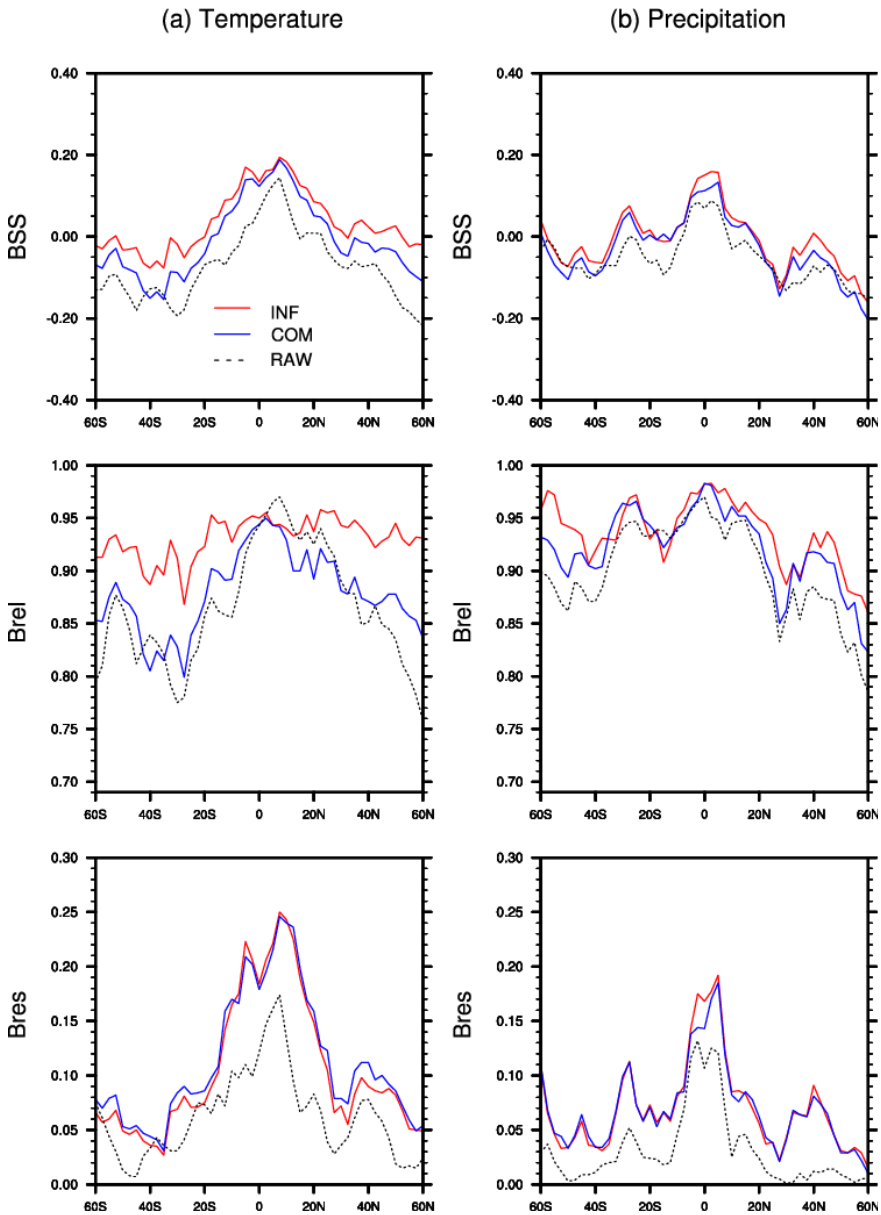

Figure 5 Zonal mean BSS of the probabilistic multi-model predictions of JJA mean (a) temperature and (b) precipitation for the period of 1981-2003 in AN categories. The blue and red solid (dashed) lines correspond to the results obtained from the SPM-multi multi-model prediction of the skill-based selected models (COM), and the inflated COM (INF), respectively. The skill of the AN categorical forecasts from the raw multi-model prediction (RAW) is also shown here with the black dashed lines, as a reference.

5. IMPROVEMENT OF THE PMMP

5.1 Application to retrospective forecast

Based on the proposed calibration and combination methods, we have developed a calibrated PMMP system and assessed its probabilistic skill in predicting AN and BN events of temperature and precipitation in comparison with the current operational version of the PMMP system (Min et al. 2009). The operational (as a control forecast) and calibrated PMMP systems (as an experimental forecast) are hereafter referred to as the 'CNT' and 'EXP', respectively.

Figures 6 and 7 show the spatial distributions of, and the differences between, the ROC scores for the AN and BN categorical forecasts of temperature and precipitation obtained from the CNT and EXP. The figures also show the results of significance tests of the estimated scores from the two PMMP systems and their differences based on the MC simulation. A comparison of the CNT and EXP clearly demonstrates that the calibrated probabilistic forecast is more skillful than the uncalibrated forecast for both categories, as the area where the estimated ROC score is statistically significant at the 5% level is considerably more extensive for the calibrated forecast. For temperature, the skill improvement given by the EXP, relative to the CNT, is generally modest than for precipitation and is mainly restricted to the tropical ocean regions. For precipitation, the significantly skillful forecast achieved by the EXP extends well into the subtropical and extratropical regions. It is interesting that the improvement of the ROC score resembles that of the TCC as shown in Fig. 3, suggesting that the positive impact of the proposed combined method is relatively larger over the tropics for temperature and over the extratropics for precipitation. This suggests that the improvement in skill of the AN and BN categorical temperature and precipitation predictions achieved by the EXP can be attributed to the calibration of model anomalies using the SPM-multi with the INF and their combination based on the skill-based selection method.

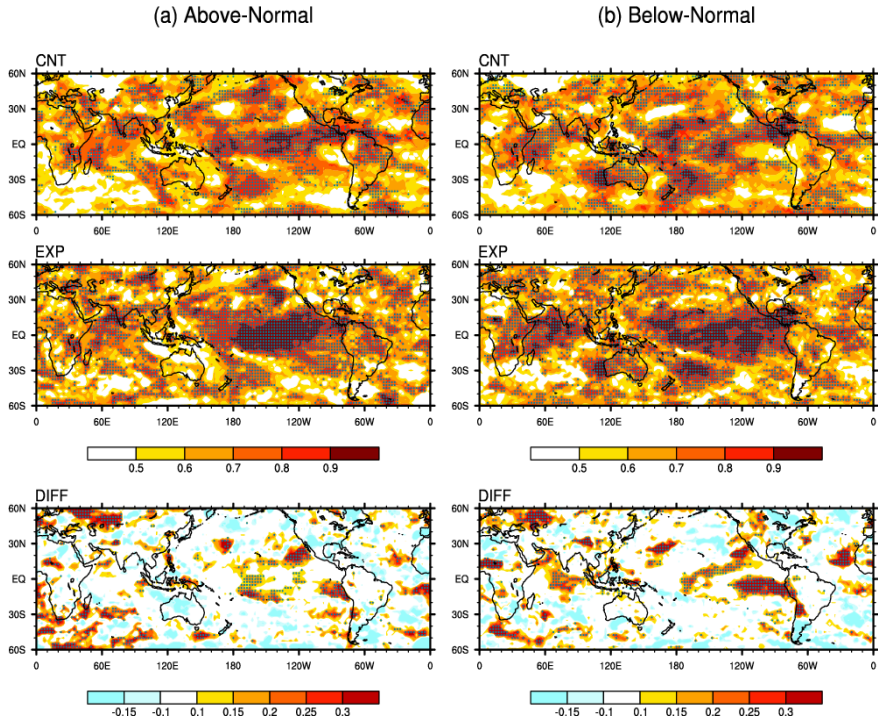

Figure 6 ROC score of the probabilistic temperature predictions in (a) AN and (b) BN categories for the period of 1981-2003, obtained from the operational version of the PMMP system (CNT) and the calibrated PMMP system (EXP). Dots indicate that the estimated ROC score from the two systems and their difference (DIFF; EXP minus CNT) are statistically significant at the 5% level using the Monte Carlo simulation.

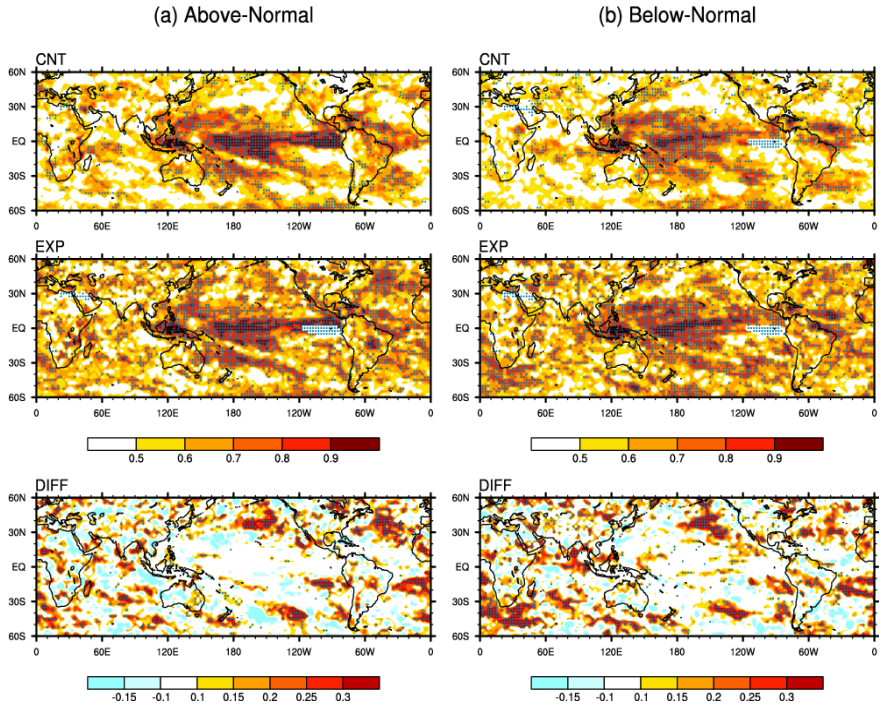

Figure 7 Same as Fig. 6 except for precipitation.

Seasonal climate prediction of the Asian summer monsoon precipitation is still limited despite the significant improvements achieved by state-of-the-art climate prediction models (Kang et al. 2002; Wang et al. 2004, 2007, 2008, 2009; Lee et al. 2010, 2011c). In this study, we use the EAM region (110°-140°E, 20°-45°N) as one of target areas to evaluate the calibrated PMMP system's capability in simulating JJA mean temperature and precipitation. For comparison, the region in which the system is most skillful (as shown in Figs. 3 and 5-7), the tropics (20°S-20°N), is also evaluated. In order to estimate large-scale verification statistics over these regions, the ROC curve and score were spatially aggregated (Fig. 8). As expected, the forecast skill of the boreal summer precipitation prediction over the EAM region is still limited in the current forecasting system, with the ROC scores of both categorical forecasts being lower than 0.6. Comparison of the ROC curves of the two systems indicates a clear improvement in the performance of the EXP over both the tropics and the EAM region, for which the ROC curve is further departed from the diagonal to the

left and upward directions than is that of the CNT. As a result, the ROC score of the calibrated system is significantly better than that of the current operational system for all aspects of the regions, variables, and categories, save only for the AN categorical temperature forecast over the EAM region, for which the improvement modesties apparent but modest. This finding implies that the EXP can discriminate between event occurrences and non-occurrences more efficiently than the CNT. The largest skill improvement of the EXP (with respect to the CNT) is commonly found for precipitation in the BN category over both the regions.

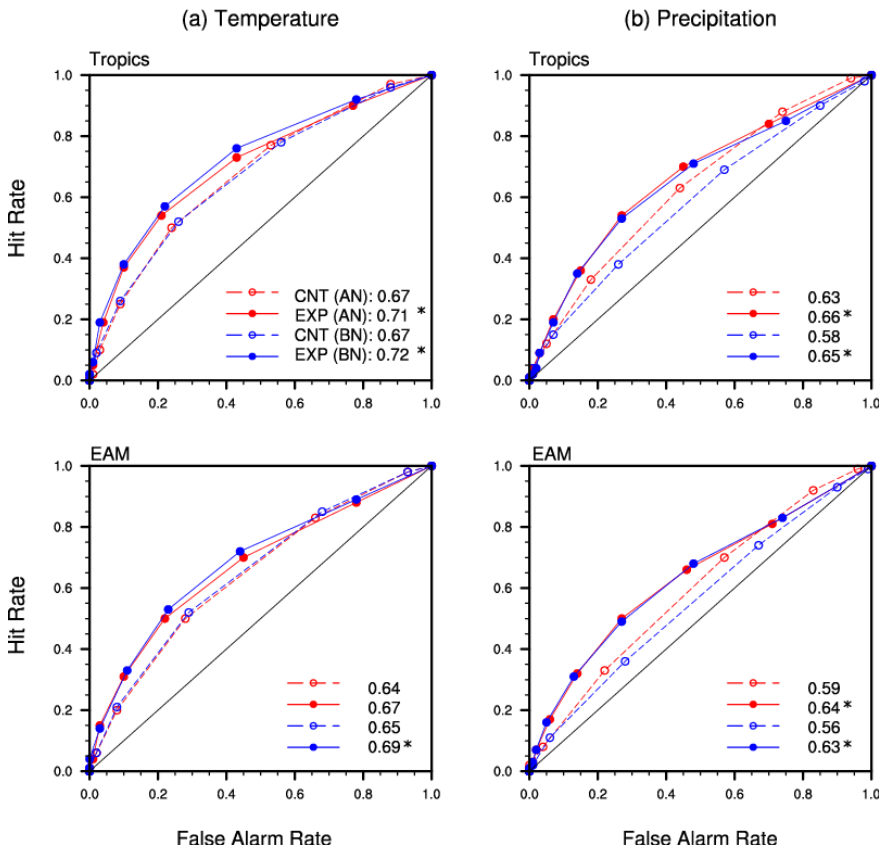

Figure 8 ROC curve of the probabilistic (a) temperature and (b) precipitation predictions in the AN and BN categories over the tropics (20°S-20°N; upper panel) and EAM region (110°-140°E, 20°-45°N; lower panel) for the period of 1981-2003. The solid (dashed) red and blue lines correspond to the results for the AN and BN categorical forecasts obtained from the EXP (CNT). The asterisk represents the skill improvement of the EXP with respect to the CNT being statistically significant at the 5% level using the Monte Carlo simulation.

For the verification of long range forecasts, the ROC score and curve used in Figs. 6-8 are more appropriate measures of forecast quality than the reliability diagram in the context of verification of long range forecast because the reliability diagram to small sample sizes (WMO 2002). However, because measures of forecast reliability are important for modelers, forecasters and end-users, the reliability diagram is recommended to use it in exceptional cases when forecasts are spatially aggregated over large regions (i.e., globe, tropics, northern and southern extratropics; WMO 2002). Following the recommendations of the WMO SVS-LRF, the reliability diagram was constructed for large-sample probabilistic forecasts aggregated over the globe to investigate the reliability of the calibrated system. Figure 9 shows reliability diagrams including unconditional frequency distribution (i.e., the sharpness; Palmer et al. 2000) of the probabilistic forecasts obtained from the CNT and EXP. In general, the reliability curves of the EXP for both variables and categories are consistently closer to the diagonal line than that of the CNT. Particularly, the forecast reliability of the calibrated temperature prediction is higher than the uncalibrated forecast, as shown by a significant increase of the reliability term of BSS (Brel). For precipitation, a general overestimation of the wide range of forecast probabilities (approx. 30%-70%) in the CNT (which leads to the shallow slope of the reliability curve and implies poor forecast resolution) can be largely corrected by the EXP, although the EXP tends to overestimate the observed frequency under high forecast probabilities (i.e., > 70%). As a result, the reliability of the two systems for precipitation, as shown by Brel, is comparable.

Additionally, the probability distribution function of the CNT for both variables and categories tends to be strongly weighted towards the climatological forecast. The results indicate that the uncalibrated probabilistic forecast is not better than a forecast based on the climatology, as shown in BSS being negative. While the calibrated system shows a significant increase in resolution term of BSS (Bres) with respect to the uncalibrated one for both temperature and precipitation. In summary, the calibrated PMMP system provides a modest improvement in reliability and a significant improvement in resolution for the AN and BN categorical temperature and precipitation forecasts over the globe for the retrospective forecast period.

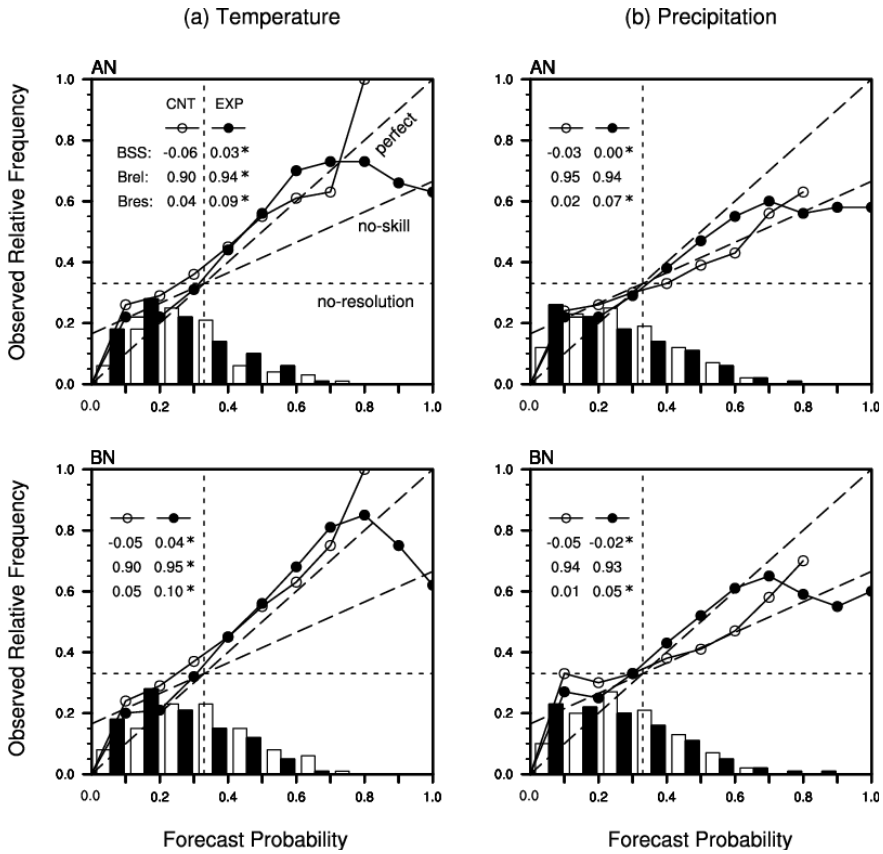

Figure 9 Reliability diagram of the probabilistic (a) temperature and (b) precipitation predictions for the AN and BN categories over the globe for the period of 1981-2003. The solid lines with closed (open) circles indicate the reliability curves of the EXP (CNT). The closed (open) bars represent the unconditional frequency distribution of the forecast (i.e., the sharpness) of the EXP (CNT). The Brier Skill Score (BSS), the reliability (Brel) and the resolution (Bres) terms of the BSS are also shown in each plot. The asterisk represents the skill improvement of the EXP with respect to the CNT being statistically significant at the 5% level using the Monte Carlo simulation.

5.2 Application to real-time forecast

To explore the usefulness of the calibrated PMMP system for real-time forecasts, which is a very important issue from an operational perspective, we have applied the proposed methods to three independent boreal summer forecasts for 2008-2010 with the training period of 1981-2003. Note that the 3-year real-time forecast period

is a common period for all of the considered climate models (Table 2), but it is not of a sufficient duration from which comprehensive conclusions can be drawn. Examples of probabilistic multi-model predictions for 2008 JJA temperature and precipitation obtained from the CNT and EXP are shown in Fig. 10, as are the corresponding observed anomalies (with respect to the period of 1981-2003). In the 2008 summer, a neutral El Niño-Southern Oscillation (ENSO) condition was observed in the equatorial Pacific, with slightly below-normal sea surface temperature (SST) over the equatorial central Pacific and above-normal SST over the eastern Pacific (APCC Outlook[10]; IRI ENSO Update/Forecast[11]). The CNT is able to capture the large-scale patterns associated with the observed ENSO phase with high probabilities of maintaining drier- and cooler–than-normal conditions over the central tropical Pacific. However, large parts of the globe are generally expected to be near-normal conditions, especially for temperature. On the other hand, the EXP provides more detailed (or localized) forecast information, although it looks quite a mosaic.

10) Information available at http://www.apcc21.org/en/services/apcc-operational-3-month-mme-prediction/forecasts/outlook/2008-son

11) Information available at http://iri.columbia.edu/climate/ENSO/currentinfo/archive/200806/update.html http://iri.columbia.edu/climate/ENSO/currentinfo/archive/200807/update.html http://iri.columbia.edu/climate/ENSO/currentinfo/archive/200808/update.html

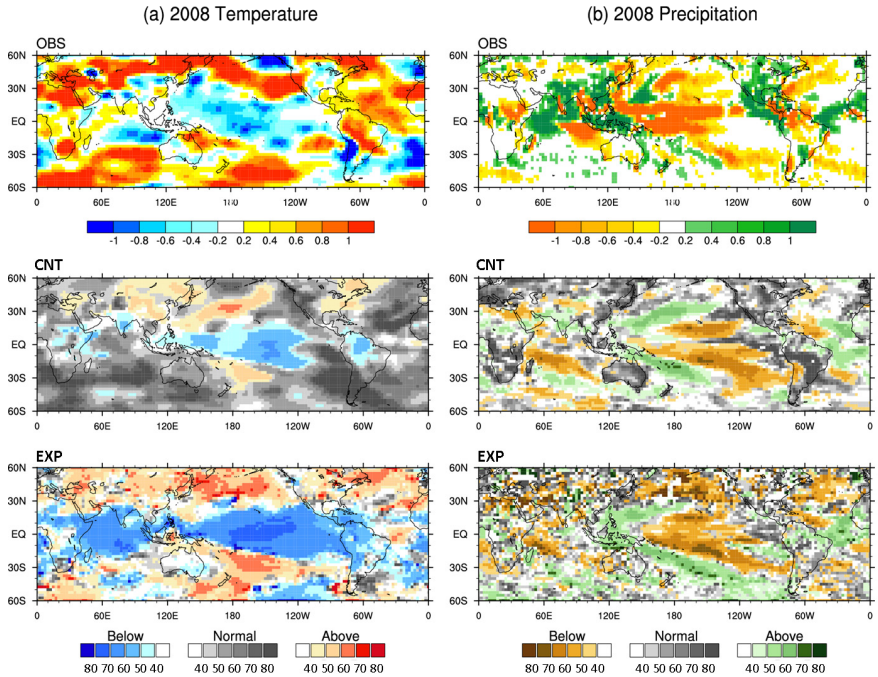

Figure 10 Observed anomalies (OBS) and predicted forecast probabilities of three events (AN, NN, and BN categories) for 2008 JJA (a) temperature and (b) precipitation by the CNT and EXP. The climatology used is for the period 1981-2003. The combined map of three categorical forecasts corresponds to the probability of the category with the highest predicted probability among the three categorical probabilities for each grid-point.

Figures 11 and 12 show the skill of the AN and BN categorical forecasts obtained from the EXP for the independent real-time forecasts (2008-2010) in terms of ROC score with respect to the CNT, along with the cross-validated retrospective forecasts (1981-2003). Also displayed are the averaged skill and the range of the highest and lowest skill scores for the whole 26-year forecast period obtained from the two systems. The positive impact of the proposed method on probabilistic forecasts can be summarized from the Figs. 11-12. First, it shows consistently better performance than the operational system not only for the retrospective forecasts but also for the real-time seasonal prediction across all aspects of variables and events. Second, its year-to-year forecast skill is more stable, indicating that the range of interannual variations in ROC score is small. Third, the skill improvement achieved for the real-time forecasts is not as large as that for the retrospective forecast. The possible reasons and potential solutions for this will be discussed in Section 6.

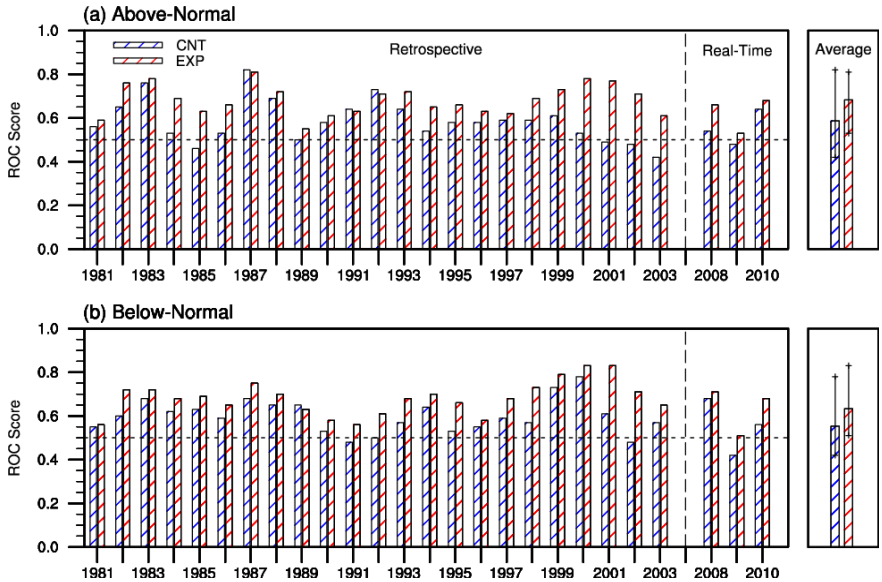

Figure 11 Time series of the ROC score of the probabilistic temperature predictions in (a) AN and (b) BN categories for the period of 1981-2003 (retrospective forecast) and 2008-2010 (real-time forecast) over the globe. The blue and red dashed bars indicate the scores of the forecast probabilities obtained from the CNT and EXP, respectively. The dashed line shows no-skill with ROC score = 0.5. The averaged scores and the range of the highest and lowest scores for the whole 26-year period of forecasts are also shown.

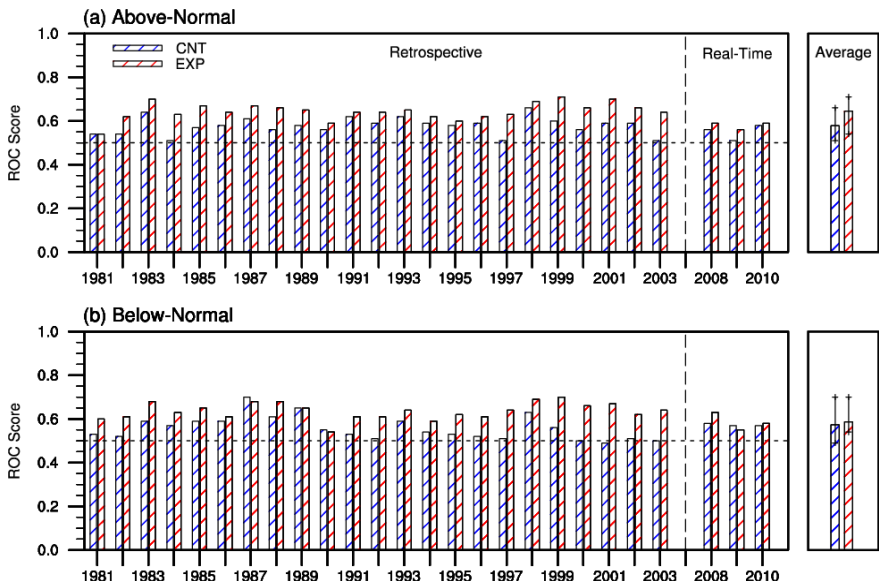

Figure 12 Same as Fig. 11 except for precipitation.

6. SUMMARY AND DISCUSSION

This study examines whether the SPM-based statistical correction method (e.g., Kang and Shukla 2006; Kug et al. 2008c) can improve the operational version of the uncalibrated PMMP system that is operationally employed at the APCC. Since previous studies have investigated the ability of the mono-variable version of the SPM in deterministic (or ensemble-mean) predictions, this study has been focused on probabilistic predictions with the upgraded multi-variable version of the SPM. In order to improve the current PMMP system, the single-model predictions have been (i) calibrated using the SPM-multi to correct the errors in predicted anomalies and variance inflation to obtain reliable probabilities and (ii) combined based on the skill-based model selection to form the multi-model prediction. Based on the proposed methods, we have developed the calibrated PMMP system and evaluated its predictive skill for the one-month lead JJA mean tercile-based categorical temperature and precipitation forecasts with respect to the uncalibrated PMMP system.

Before trying to develop the calibrated PMMP system, a comprehensive assessment of the impact of model calibration (using the SPM-multi and variance inflation) and combination (based on the skill-based model selection) is conducted based on the retrospective forecasts for the period of 1981-2003. The single-model and multi-model predictions have been corrected using the previous and upgraded versions of the SPM, and the forecast skills of the prediction methods have been compared with the raw single-model outputs and their simple averaged MME (with equal weighting). The results indicate that the corrected single-model and multi-model predictions using both versions of the SPM clearly show large improvements in TCC relative to the raw predictions. The positive impact of the single-model correction is larger than that of the simple multi-model correction, particularly for the models whose raw predictions are particularly unskillful. A comparison of the different versions of the SPM indicates that the SPM-multi has consistently better performance than the SPM-mono when considering all aspects of the single-model and the simple multi-model for both variables.

The corrected single-model predictions using both versions of the SPM have been combined based on the skill-based model selection in the formulation of multi-model predictions. The results indicate that model combination is also an important factor in improving the forecast quality of the corrected multi-model predictions by excluding less skillful model's contributions to the MME with regards to the simple multi-model prediction. Model combination is found to improve both versions of the SPM and variables, and its benefit is relatively larger for precipitation (than temperature) and the SPM-multi (than the SPM-mono). As a result, the largest improvements due to model combination are found for the corrected multi-model precipitation predictions using the SPM-multi.

For a regionalized assessment of the impact of the proposed methods (the SPM-multi with variance inflation and model combination), the zonally averaged TCC and BSS were used. In general, both model calibration and combination increase the skill of temperature and precipitation predictions over most latitudinal zones. The benefit of model combination is relatively larger over the extratropical regions for both temperature and precipitation, while the skill increase due to model correction is larger over the tropical region for temperature and consistently observed over all the latitudes for precipitation. As a result, the most beneficial impact of the methods, relative to the raw prediction, is found when the proposed correction and combination methods are combined. That is, the combined approach is significantly superior to the separate methods for both variables over most of the latitudes. To adjust (or rescale) the variance of the corrected predictions to match the corresponding observed variance for accurate estimation of forecast uncertainty, an additional calibration, variance inflation, has been applied. It is shown that variance inflation of the corrected single-model predictions can contribute not only to correcting the underestimation of the predicted IAV, but also to accurately estimating forecast uncertainty, which is important for quantifying the skill in a probabilistic framework. Variance inflation is shown to increase the BSS over most of the latitudes, primarily due to the associated increase in reliability, particularly for temperature over the extratropical region.

Finally, we have assessed the forecast skill obtained with the calibrated PMMP system, which has been improved by the proposed calibration and combination

methods, in comparison with the current operational version of PMMP, based on the retrospective (1981-2003) and real-time forecasts (2008-2010). The results indicate that the calibrated probabilistic forecasts are more skillful than the uncalibrated forecasts, especially over the tropical (extratropical) regions for temperature (precipitation), and as is not the case for the current operational PMMP, the significantly skillful forecasts are extended well into the extratropical regions. Moreover, the proposed approach improves the PMMP forecast skill for the boreal summer precipitation over the EAM region, which is one of the challenges in current climate system. While the increase in resolution by the calibrated probabilistic forecasts is significant, the improvement in reliability is modest. Comparison of the operational and calibrated PMMP systems also revealed that the calibrated probabilistic forecasts show consistently better (with higher ROC score) and more stable performance (with smaller interannual variations of ROC score) than the uncalibrated forecasts across all variables and categories, throughout the whole 26-year period of retrospective and real-time forecasts.

As mentioned in Section 3.1, we have adopted the SPM-based correction method. The novelty of the method is to objectively and automatically select the predictor domain based on its statistical relationship with the predictand during the cross-validated training period. The method differs from the most widely used statistical models, which have been developed with fixed predictor domains determined from empirical relationships, in that achieves a high level of skill by implicitly considering only significant relationships (e.g., Kang et al. 2007; Zhu et al. 2008; Chu et al. 2008). However, there might be some difficulties in applying this fixed predictor domain to operational processes for real-time forecasts. Especially, prediction skill might be lower for real-time forecasts than for the training period due to the so-called overfitting problem, as discussed in Kug et al. 2008b. In this regard, the objective and automatic selection of the predictor domain is appropriate for use in operational prediction system and it has a lot of potential to change the predictand or extend the number of predictors, as shown in the study with increased number of potential predictors.

In the present study, we have described the positive impact of the proposed

calibration and combination methods in both deterministic and probabilistic prediction. Results indicate that both methods increase the forecast skill for both retrospective and real-time forecasts for JJA mean temperature and precipitation. However, as shown in Figs. 11-12, the increased skill (with respect to the uncalibrated PMMP system) for the real-time forecasts is not as large as that for the retrospective forecasts. One of the reasons might be that the 23-year training period used in the study, which is a common period for all of the climate models, is not sufficient to make stable relations in the SPM-based statistical correction model. In a statistical model, it is important to capture the stable and strong relationships between predictor(s) and predictand in the long-training period and the statistical correction models are sensitive to the training period (e.g., Kug et al. 2008b; Doblas-Reyes et al. 2005). Another reason for the real-time forecasts being less improved than the retrospective forecasts might be the discontinuity in the training and real-time forecast periods, that is, the recent climate pattern of 2004-2007 was not reflected in the statistical model due to its not being component to the hindcast dataset employed by the climate models. In this regard, there is substantial room for further improvement of the predictive skill of the proposed statistical correction method by increasing the duration of the training period.

Appendix A

As mentioned in Section 3, this study uses the upgraded multi-variable version of the SPM which includes as one of the calibration methods an increased number of potential predictors to correct the errors in model anomalies. To select the additional potential predictors among the 8 variables which are common among all the models (Table 2), we explored their skills in terms of TCC with respect to the observations for the period of 1981-2003. The CAMS OPI and NCEP-DOE reanalysis 2 were used as verification data for precipitation and other variables, respectively. Figure A1 shows the prediction skills of the individual models and their simple averaged MME with equal weighting over the globe (60oS-60oN) for each of the potential predictors; temperature at 850hPa (T850), precipitation (PREC), geopotential height at 500hPa (Z500), sea level pressure (SLP), and wind fields at 850hPa and 200hPa (U/V200 and U/V850). The results indicate that the statistically significant skills at the 10% level using the one-tailed t-test are found for Z500 and SLP, thus they are used in the study as additional predictors.

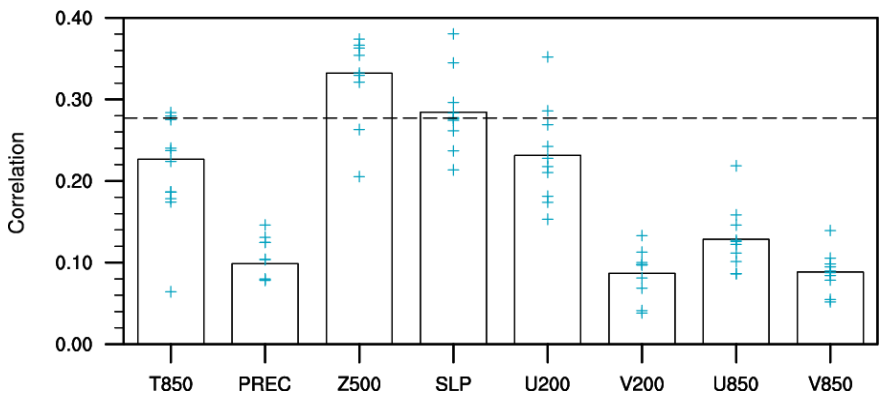

Figure A1 TCC of ten dynamical model predictions and their simple averaged multi-model prediction for each potential predictor, during the 23-year hindcast period of 1981-2003. The open bars and cross markers represent the skills of the simple averaged multi-model with equal weighting and the individual single-model predictions, respectively. The dashed line represents the TCC value being statistically significant at the 10% level using the one-tailed t-student test.

REFERENCES

Alessandri A, Borrelli A, Navarra A, Arribas A, Deque M et al. (2011) Evaluation of probabilistic quality and value of the ENSEMBELS multi-model seasonal forecasts: comparison with DEMETER. Mon Weather Rev 139:581-607. doi:10.1175/2010MWR3417.1

Atger F (2003) Spatial and interannual variability of the reliability of ensemble-based probabilistic forecasts: Consequences for calibration. Mon Weather Rev 131:1509-1523

Back SK, Ryu JH, Ryoo SB (2002) Analysis of the CO2 doubling experiment using METRI AGCM. Part I: The characteristics of regional and seasonal climate responses. J Korean Meteor Soc 38: 465–477

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

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

Chu JL, Kang H, Tam CY, Park CK, Chen C (2008) Seasonal forecast for local precipitation over northern Taiwan using statistical downscaling. J Geophys Res 113:D12118. doi:10.1029/2007JD009424

Coelho C.A.S., D. B. Stephenson, M. Balmaseda, F. J. Doblas-Reyes and G. J. van Oldenborgh (2006) Towards an integrated seasonal forecasting system for South America. J Clim 19:3704-3721

Cote J, Gravel S, Méthot A, Patoine A, Roch M, Staniforth A (1998) The operational CMC/MRB global environmental multiscale (GEM) model: Part I - Design considerations and Formulation. Mon Weather Rev 126: 1373-1395

Cusack S, Arribas A (2009) Sampling errors in seasonal forecasting. Mon Weather Rev 137:1132-1141

Derome J et al (2001) Seasonal predictions based on two dynamic models. Atmos-Ocean 43:129-143

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

Hamill T, Colucci SJ (1998) Verificaiton of Eta-RSM ensemble probabilistic precipitation forecasts. Mon Weather Rev 126: 711-724

Huth R (1999) Statistical downscaling in central Europe: Evaluation of methods and potential predictors. Clim Res 13: 91-101

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 (2011) Forecast skill for the leading seasonal atmospheric pattern coupled with the tropical Pacific SST. Will be submitted to Clim Dyn

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

Jolliffe IT, Stephenson DB (2003) Forecast Verification: A Practitioner's Guide in Atmospheric Science. John Wiley and Sons, UK, 240 pp

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, Ah KH, Park CK et al (2007) Multimodel output statistical downscaling prediction of precipitation in the Philippines and Thailand. Geophys Res Lett 34 L15170. doi:10.1029/2007GL030730

Kang IS et al (2002) Intercomparison of the climatological variations of Asian summer monsoon precipitation simulated by 10 GCMs. Clim Dyn 19:383-395. doi:10.1007/s00382-002-0245-9

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

Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D (1990) A method of relating general circulation model simulated climate to observed local climate. Part I: Seasonal statistics. J Clim 3:1053-1079

Ke Z, Zhang P, Dong W, Le L (2009) A new way to improve seasonal prediction by diagnosing and correcting the intermodel systematic errors. Mon Weather Rev 137:1898-1907

Kharin VV, Zwiers FW (2001) Skill as function of time scale in ensemble of seasonal hindcast. Clim Dyn 17:127-141. doi: 10.1007/s003820000102

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

Kharin VV, Zwiers FW (2003b) On the roc score of probability forecasts. J Clim 16: 4145-4150.

Klen WH, Lewis BM, Enger I (1959) Objective prediction of five-day mean temperatures during winter. J Meteorol 16: 672-682

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〉2.0CO;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

Krzysztofowicz R (1983) Why should a forecaster and a decision maker use Bayes theorem. Water Resour Res 19:317-336

Kug JS, Kang IS, Choi DH (2008a) Seasonal climate predictability with tier-one and tier-two prediction systems. Clim Dyn 31:403-416. doi:10.1007/s00382-007-0264-7

Kug JS, Lee JY, Kang IS (2007) Global sea surface temperature prediction using a multimodel ensemble. Mon Weather Rev 135:3239-3247

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

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, Wang B, Ding Q, Ha JY, Ahn JB et al (2011b) How predictable is the northern hemisphere summer upper-tropospheric circulation? Clim Dyn 37:1189-1203

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, SS, Lee JY, Ha KJ, Wang B, Schemm JKE (2011c) Deficiencies and possibilities for long-lead coupled climate prediction of the Western North Pacific-East Asian summer monsoon. Clim Dyn 36:1173-1188

Liou CS, Chen JH, 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: 653–663

Manson I (1982) A model for assessment of weather forecasts. Aust Meteor Mag 30:291-303

McFarlane N, Boer GJ, Blanchet JP, Lazare M (1992) The Canadian climate centre second generation general circulation model and its equilibrium climate. J Clim 5: 1013–1044

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

Murphy AH (1973) A new vector partition of the probability score. J Appl Meteor 12:595-600

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

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

Park CK et al (2002) Long-term forecasting system. Climate Science Bureau Tech. report no. 2002-8, Korea Meteorological Administration, Seoul

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

Richardson DS (2000) Skill and economic value of the ECMWF ensemble prediction system. Q J R Meteorol Soc 126:649-668

Ritchie H (1991) Application of the semi-Lagrangian method to a multilevel spectral primitive-equations model. Q J R Meteor Soc 117: 91-106

Saha S, Nadiga S, Thiaw C, Wang J et al (2006) The NCEP climate forecast system. J Clim 19: 3483–3517

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

Stefanova L and Krishnamuri TN (2002) Interpretation of seasonal climate forecast using Brier skill score, the Florida State University superensemble, and the AMIP-I data set. J Clim 15:537-544

Stephenson DB, Doblas-Reyes FJ (2000) Statistical methods for interpreting Monte Carlo ensemble

forecasts. Tellus 52A:300-322

Taylor JR (1982) An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, University Science Books, 327 pp

Tippett MK, Barnston AG, Robertson AW (2007) Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. J Clim 20:2210-2228

van den Dool H, Toth Z (1991) Why do forecast for "near normal" often fail? Wea Forecasting 6:76-85

von Storch H (1999) On the use of "inflation" in statistical downscaling. J Clim 12:3505-3506

von Storch H, Zwiers FW (1999) Statistical Analysis in Climate Research, Cambridge Univ Press, New York, 484 pp

Wang B, Kang IS, Lee JY (2004) Ensemble simulations of Asian-Australian monsoon variability by 11 AGCMs. J Clim 17:803-818

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 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, Lee JY, Kang IS, Shukla J, Saji NH, Pack CK (2007) Coupled predictability of seasonal tropical precipitation. CLIVAR Exchanges 12:17-18

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.

Wetterhall F, Halldin S, Xu CY (2005) Statistical precipitation downscaling in central Sweden with the Analogue Method. J Hydrol 306:174-190. doi:10.1016/j.jhydrol.2004.09.008

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

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

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

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

Zhu C, Park CK, Lee WS, Yun WT (2008) Statistical downscaling for multi-model ensemble prediction of summer monsoon rainfall in the Asia-Pacific region using geopotential height field. Adv Atmos

Sci 25:867-884. doi:10.1107/s00376-008-0867-x

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

Zwiers FW (1996) Interannual variability and predictability in an ensemble of AMIP climate simulations conducted with the CCC GCM2. Clim Dyn 12: 825-848. doi: 10.1007/s003820050146

APCC TECHNICAL REPORT 2011-02

- Improvement of MME Seasonal Prediction Skill
- Assessment of the Long-Lead Probabilistic Prediction
- Improvement of the APCC Probabilistic Multi-Model Ensemble Prediction

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 533172
ISBN 978-89-97333-17-2
ISBN 978-89-97333-15-8 (세트)