RESEARCH ARTICLE

Assessment of MME methods for seasonal prediction using WMO LC-LRFMME hindcast dataset

Gayoung Kim[1,2] | Joong-Bae Ahn[2] ⓘ | Vladimir N. Kryjov[2] | Woo-Seop Lee[1] |
Dong-Joon Kim[3] | Arun Kumar[4]

[1]APEC Climate Center, Busan, South Korea

[2]Department of Atmospheric Sciences, Pusan National University, Busan, South Korea

[3]Korea Meteorological Administration, Seoul, South Korea

[4]Climate Prediction Center, National Centers for Environmental Prediction, College Park, Maryland

Correspondence
Joong-Bae Ahn, Department of Atmospheric Sciences, Pusan National University, Geumjeong-gu, Busan 46241, Republic of Korea.
Email: jbahn@pusan.ac.kr

Abstract

Different multi-model ensemble (MME) methods were investigated for their potential to improve the skill of 1-month lead seasonal forecast products, based on six models from Global Producing Centers (GPCs) for long-range forecasts (LRFs) designated by the World Meteorological Organization (WMO). We first compared the hindcast performance of seven MME methods (simple composite method, SCM; simple linear regression, SLR; multiple linear regression, MLR; best selection anomaly, BSA; multilayer perceptron, MLP; radial basis function, RBF; genetic algorithm, GA) for the global 2-m temperature and precipitation for 1983–2009. The reference datasets for 2-m temperature and precipitation are the ERA-Interim from European Centre for Medium-Range Weather Forecasts (ECMWF) and Global Precipitation Climatology Project (GPCP) for hindcast verification. For real-time verification, the data from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis 1 for 2-m temperature and climate anomaly monitoring system and outgoing longwave radiation precipitation index (CAMS OPI) for precipitation are used. According to our analysis, GA was the most successful MME method in predicting both the global 2-m temperature and precipitation for all four seasons. GA also showed good performance in predicting the 2-m temperature and precipitation over the 13 regional climate outlook forum (RCOF) regions in all four seasons, but the range in performance among the RCOF regions varied significantly. In a real-time forecast period (MAM 2012-DJF 2015/16), GA outperformed in terms of time-averaged anomaly pattern correlation coefficient (ACC) and root-mean-square error (RMSE) of the 2-m temperature, although the forecast skill difference (0.02) between GA and SCM was not statistically significant. For the precipitation, both SCM and GA also reveal better performance than other MME methods. During the very strong El Niño event in 2015, individual models show better performance than other years. Nonetheless, these two MME methods outperform all the individual models.

1 | INTRODUCTION

Dynamical predictions on a seasonal timescale, which are based on modelling the physical processes in the atmosphere, ocean, cryosphere, and land surface, are a viable tool for anticipating the near-time evolution of the climate system. However, the skill of dynamical seasonal predictions has remained limited, partially because of systematic errors in models arising from factors such as model parameterization schemes (Zadra et al., 2018), and the model's inability to produce an atmospheric response to oceanic boundary conditions (Kang et al., 2004; Park et al., 2018).

Over the last two decades, multi-model ensemble (MME) approaches that combine forecasts from different climate models have been introduced to reduce the influence of errors associated with model biases (e.g., Krishnamurti et al., 1999, 2000; Shukla et al., 2000; Peng et al., 2002; Palmer et al., 2004; Hagedorn et al., 2005; Wang et al., 2009; Mishra et al., 2019). The equal-weighed MME method generally shows consistently better prediction skill each individual model does because the errors from individual models tend to offset each other (Kharin and Zwiers, 2002; Peng et al., 2002; Hagedorn et al., 2005; DelSole, 2007). This is particularly important in long-term forecasts because errors in the model are gradually accumulated and increased. Meanwhile, unequal-weighted MME forecasts that give more weights to the models with consistently better performance have also been suggested (Krishnamurti et al., 2000; Yun and Krishnamurti, 2002; Ahn and Lee, 2016; Mishra et al., 2019). Many operational centres for seasonal prediction have now adopted the MME approach to provide more reliable forecasts to users (Palmer et al., 2004; Kirtman et al., 2014; Min et al., 2014, 2017; Sohn et al., 2016; Kim et al., 2016). Recent MME studies have expanded its scope to study extreme weather conditions such as droughts, typhoons, monsoons and heat waves using seasonal forecast data (Sohn et al., 2011; Sohn and Tam, 2016; Kim et al., 2017; Kim, 2018; Kim and Chan, 2018; Park et al., 2018; Shin and Moon, 2018).

The World Meteorological Organization Lead Center for Long-Range Forecast MME (WMO LC-LRFMME), established in 2009 and jointly operated by the Korea Meteorological Administration (KMA) and the National Oceanic and Atmospheric Administration (NOAA)/NCEP, provides a conduit for sharing model forecasts from global producing centres (GPCs) for long-range forecasts (LRFs).

The LC-LRFMME was initiated to provide a well-calibrated MME system and a "one-stop shop" for GPC information (Graham et al., 2011; Kim et al., 2016). The WMO LC-LRFMME issues global probabilistic and deterministic seasonal forecasts that are openly available via its website (https://www.wmolc.org/).

In this study, we evaluate the methods for creating deterministic forecasts, that is, forecasts of the MME anomalies, by combining forecasts from GPC-LRFs. The study results are important for improving probabilistic forecasts, as well. The WMO LC-LRFMME provides the probabilistic forecasts using the probabilistic MME method developed by Min et al. (2009) and currently implemented at the APEC Climate Center (APCC). In this method, predicted probabilities are approximated with a Gaussian probability distribution function (PDF). The first parameter of the predicted Gaussian PDF, the mean, can be improved based on the results from this study, while the second parameter, the variance, can be assessed by accounting for both ensemble spread and calibration error as has been demonstrated for Korea by Min et al. (2011) and for Northern Eurasia by Kryzhov (2012).

Besides WMO LC-LRFMME, various deterministic MME methods are used operationally in several operational centres, particularly APCC, Copernicus Climate Change Services (C3S), and North American MME (NMME). Since its establishment in 2005, APCC has issued monthly 3-month global predictions of the 2-m temperature and precipitation. APCC uses four deterministic MME methods (simple composite mean, SCM; stepwise pattern projection method, SPM; multiple linear regression, MLR; synthetic superensemble method, SSE) for combining the ensemble means of participating models that are operationally exploited for seasonal forecasts using three Atmospheric General Circulation Models (AGCMs) and 10 Coupled General Circulation Models (CGCMs) from leading climate forecasting centres and institutes in 10 countries (Kryjov et al., 2006; Kirtman et al., 2014). C3S was implemented by the European Center for Medium-Range Weather Forecasts (ECMWF) as part of the delegation agreement with the European Union and publishes seasonal forecast products every month. Current participating centres in C3S are ECMWF, The Met Office, Météo-France, the German Meteorological Service (Deutscher Wetterdienst, DWD), and the Euro-Mediterranean Center on Climate Change (Centro Euro-Mediterraneo sui Cambiamenti Climatici,

CMCC). Simple-averaged MME with equal weighting is used with a correction for bias and variance estimated from the full set of hindcasts available for each model, which is called the "standard bias correction technique" (Stockdale, 2012). NMME is a real-time experimental operational forecast system consisting of coupled models from North American modelling centres, including NOAA/NCEP, NOAA/Geophysical Fluid Dynamics Laboratory (GFDL), International Research Institute for Climate and Society (IRI), National Center for Atmospheric Research (NCAR), National Aeronautics and Space Administration (NASA), and Canadian Meteorological Center (CMC), starting in August 2011. Among other methods, NMME also uses simple-averaged MME with anomaly forecasts for each model based on the climatology from the hindcasts (Kirtman *et al.*, 2014).

WMO LC-LRFMME provides four deterministic MME predictions (SCM; regular multiple regression, MLR; singular value decomposition, SVD; genetic algorithm, GA) for a 1-month lead seasonal forecast provided by 13 WMO-designated GPCs (Kim *et al.*, 2016). Several years have passed since WMO LC-LRFMME started providing MME forecasts; nevertheless, the skill of MME methods using GPCs of WMO LC-LRFMME has not been assessed in an operational context.

Although many MME techniques for seasonal prediction have been developed for research and/or operational purposes, the MME techniques were evaluated individually, sporadically and in isolation. The purpose of this study is to assess an overall performance of the several MME methods using the world's leading models for seasonal forecasting. We first introduce the various deterministic MME methods. Next, we assess the general performance of different MME predictions with the bias-corrected ensemble mean forecast worldwide and then highlight the MME method with the highest prediction skill in the hindcast period. We also examine the performance of the selected MME over 13 regional climate outlook forum (RCOF) regions for the hindcast period. Finally, we apply the MME method to real-time forecasting and present the prediction skills of the MME.

2 | DATA AND METHOD

2.1 | Data and verification metrics

The MME predictions using six models from five GPCs participating in WMO LC-LRFMME are used (Table 1). The other GPC hindcast data of which the periods are not long enough (e.g., ECMWF and Toulouse) or do not match others (e.g., Beijing, CPTEC, Exeter, Pretoria, Seoul, and Offenbach) are not included in the analysis. Moscow is the only uncoupled model, while the others are coupled models (Melbourne, Montreal [CANCM3 and CANCM4], Tokyo, and Washington). The hindcasts cover the 27 years from 1983 to 2009, and their real-time forecasts from 2012 to 2015 (12 independent predictions) are analysed. The 1-month lead retrospective and real-time MME forecasts for the seasons of March–April–May (MAM) initialized on February, June–July–August (JJA) initialized on May, September–October–December (SON) initialized on August, and December–January–February (DJF) initialized on November are evaluated. All prediction skills are focused on the 2-m temperature and precipitation. The datasets are interpolated into a grid with a resolution of 2.5° in both longitude and latitude. The

TABLE 1 WMO GPCs whose predictions are used in this study with summarized information for the forecast system configurations

GPC name (model)	Organization	Resolution (atmospheric component)	Hindcast period	System configuration (ensemble size of hindcast)
Melbourne (POAMA M2.4)	Australian Bureau of Meteorology	T47/L17	1980–2010	Coupled (99)
Montreal (MSC CANCM3)	Meteorological Service of Canada	T63	1981–2010	Coupled (10)
Montreal (MSC CANCM4)	Meteorological Service of Canada	T63	1981–2010	Coupled (10)
Moscow (SL-AV)	Hydrometeorological Centre of Russia	1.125° × 1.40625°/L28	1981–2010	Uncoupled (10)
Tokyo (JMA/MRI-CPS2)	Japan Meteorological Agency	1.875° × 1.875°/L40	1979–2014	Coupled (10)
Washington (NCEP CFSv2)	National Centers for Environmental Prediction	T126/L64	1983–2009	Coupled (20)

Note: More details are available at WMO LC-LRFMME website (www.wmolc.org).

anomalies are calculated as departures from the climatology of each model over the training period in the leave-one-out cross-validation scheme to provide comparable results (Wilks, 1995). All of the non-equal-weights combination methods are calculated in the leave-one-out cross-validation mode.

Following official recommendations of the WMO Joint CBS-CCl Expert Team on Operational Predictions from Sub-seasonal to Longer-time Scales (ET-OPSLS) adopted at the meeting held on 11–15 April 2016 in Beijing, China, we use different datasets for hindcast and real-time forecast verification. The data used for verification of the retrospective forecast are 2-m temperature from the ERA-Interim reanalysis (Dee *et al.*, 2011) of European Centre for Medium-Range Weather Forecasts (ECMWF) and precipitation from the Global Precipitation Climatology Project (GPCP) merging data from rain gauge stations, satellites, and sounding observation (Adler *et al.*, 2003). Verification of the real-time forecasts was performed on the near real-time 2-m temperature from NCEP/NCAR reanalysis 1 (Kalnay *et al.*, 1996) which is a joint product from NCEP and NCAR and precipitation from Climate Anomaly Monitoring System and Outgoing longwave radiation Precipitation Index (CAMS OPI) merging observation from rain gauges with satellite estimates to obtain the quasi real-time monthly mean global precipitation dataset (Janowiak and Xie, 1999). The anomalies from observation are also leave-one-out cross-validated data (e.g., WMO, 2010; Lee *et al.*, 2011, 2013a; Min *et al.*, 2014, 2017; Mishra *et al.*, 2019). The Niño 3.4 index based on optimum interpolation (OI) version 2 monthly mean sea surface temperature (SST) (Reynolds *et al.*, 2002) obtained from the National Oceanographic and Atmospheric Administration (NOAA)/Climate Diagnostics Center (CPC) (available at https://www.cpc.ncep.noaa.gov/data/indices/) is used to present the relationship between forecast skill and El Niño-Southern Oscillation (ENSO) variability. OISST version 2 was the officially suggested dataset for verifying sea surface temperature at the meeting of CBS for Expert Team on Develop a Verification System for Long-range Forecasts held on 16–20 August 1999 in Reading, United Kingdom (the final report is available at: https://www.wmo.int/pages/prog/www/DPS/Reports/ET-LRF-DevVS_ecmwf1999.html).

The seasonal forecast skill of the various MME methods is assessed using the *SD*, the temporal correlation coefficient (TCC), the anomaly pattern correlation coefficient (ACC), and the root-mean-square error (RMSE). The *SD*, TCC, and RMSE are calculated for each grid point and then globally averaged. We also estimate the time-averaged ACC and RMSE worldwide. The scores are defined as follows:

$$TCC_j = \frac{\sum_{i=1}^{N} F'_j O'_j}{\sqrt{\sum_{i=1}^{N} F'^2_j} \sqrt{\sum_{i=1}^{N} O'^2_j}} \quad (1)$$

$$ACC = \frac{\sum_{j=1}^{n} \cos\theta_j F'_j O'_j}{\sqrt{\sum_{j=1}^{n} \cos\theta_j F'^2_j} \sqrt{\sum_{j=1}^{n} \cos\theta_j O'^2_i}} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(F_j - O_j\right)^2}{n}} \quad (3)$$

where N is the number of years of hindcast period, n is the number of grids, F and O denote forecast and observation and F' and O' are the anomaly for forecast and observation, respectively. j denotes each grid point.

According to the agreements on the need to have a more coherent approach to verification of long-range forecast, the Standard Verification System (SVS) for LRF (SVS-LRF) has arranged the procedures for the practical details of producing and exchanging appropriate verification scores (WMO, 2010). The mean square skill score (MSSS) is recommended by SVS-LRF for deterministic verification. The MSSS is based on the mean squared error (MSE) between a set of paired hindcast (MSE_j) and observations ($MSEc_j$).

$$MSSS_j = \frac{MSE_{cj} - MSE_j}{MSE_{cj}} = 1 - \frac{MSE_j}{MSE_{cj}} \quad (4)$$

$$MSE_j = \frac{1}{N} \sum_{i=1}^{N} \left(F_{ij} - O_{ij}\right)^2 \quad (5)$$

$$MSEc_j = \left(\frac{N}{N-1}\right)^2 \frac{1}{N} \sum_{i=1}^{N} \left(F_{ij} - \bar{F}_j\right)^2 \quad (6)$$

The overall MSSS is computed as below:

$$MSSS = 1 - \frac{\sum_{j=1}^{n} \cos\theta_j MSE_j}{\sum_{j=1}^{n} \cos\theta_j MSE_{cj}} \quad (7)$$

MSSS can be expanded (Murphy, 1998) as.

$$MSSS_j = \left\{ 2\frac{S_{F_j}}{S_{O_j}} r_{O_j} - \left(\frac{S_{F_j}}{S_{O_j}}\right)^2 - \left(\frac{\left[\bar{F}_j - F\bar{O}_j\right]^2}{S_{O_j}}\right) + \frac{2N-1}{(N-1)^2} \right\} /$$
$$\left\{ 1 + \frac{2N-1}{(N-1)^2} \right\}$$
$$(8)$$

where r_{FO_j} is the product moment correlation of the forecasts and observations at point j as

$$r_{FO_j} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(F - \bar{F}_j\right)\left(O_{ij} - \bar{O}_j\right)}{S_{Fj}S_{Oj}} \qquad (9)$$

Here, S_{F_j} and S_{O_j} are the SD of forecast and observation, respectively. The first three terms of the decomposition of $MSSS_j$ are related to phase errors (through the correlation), amplitude errors (through the ratio of the forecast to observed variances) and overall bias error, respectively, of the forecast. These values can be helped to adjust or weight the forecasts used as inputs to regional and local forecasts Good model performance with strong correlation and small amplitude error can be considered suitable for use in prediction.

2.2 | MME prediction methods

We suggest seven experimental deterministic MME methods to merge the six seasonal forecast systems. Each method is described below and summarized in Table 1.

2.2.1 | Simple composite method (SCM)

SCM is the simplest and most commonly used way to combine the bias-corrected seasonal forecast with equal weights (see Equation (10)), where F'_i is the forecast anomaly of the ith model, and N is the number of individual models involved in MME. SCM is widely used in many projects or operational centres (e.g., APCC, DEMETER, NMME, WMO LC-LRFMME), as recommended by WMO (Palmer et al., 2004; WMO, 2007, 2008; Kirtman et al., 2014; Kim et al., 2016; Min et al., 2017).

$$MME = \frac{1}{N}\sum_{i=1}^{N}F'_i \qquad (10)$$

2.2.2 | Linear regression methods

This approach uses empirically weighted MME with coefficients calculated for simple linear regression (SLR) and MLR. SLR is given by Equation (11), where a_i is the regression coefficient of the ith model.

$$MME = \frac{1}{N}\sum_{i=1}^{N}a_iF'_i \qquad (11)$$

MLR is built using covariance matrices to minimize the mean-square error during the training period and is calculated by

$$MME = \sum_{i=1}^{N}b_iF'_i \qquad (12)$$

where b_i is the MLR coefficient of the ith model, which is obtained by minimizing the difference between the individual models and observations (Krishnamurti et al., 1999, 2000; Yun et al., 2003; Ke et al., 2009).

2.2.3 | The best selection anomaly (BSA) method

The best selection anomaly (BSA) method is a linear method to merge models. This method takes the forecasted anomaly of the ith model $\left(Fz_{ij} - \bar{F}_{ij}\right)$, which has the highest correlation with observations among the individual models for the jth point in the cross-validation process (KMA, 2009a, 2009b).

$$MME_j = F'_{ij} \qquad (13)$$

2.2.4 | Artificial neural network (ANN) method

The artificial neural network (ANN) is a nonlinear MME prediction method that is used for nonlinear regression and its classification problems. Previous studies demonstrate that the seasonal prediction estimated by ANN outperforms that of each participating model, and that it is more advantageous than MLR in its ability to extract useful information in the case of nonlinear cause–effect relationships (Yun and Krishnamurti, 2002; Park et al., 2005; Ahn et al., 2012). ANN consists of three types of layer: input, hidden, and output. The forecast data are entered into the input layer, and the MME results are generated from the output layer. There is more than one hidden layer between the input and output layers (see figure 6 in Yun and Krishnamurti, 2002).

We choose the multilayer perceptron (MLP) algorithm, which is the most widely used algorithm to calculate optimal weighting (Marius-Constantin et al., 2009). The jth neuron in the hidden layer is assigned the value, y_j,

$$y_j = \tanh \sum_i^N \left(w_{ij} x_i + c_j \right) \tag{14}$$

where N is the number of hidden neurons, w_{ij} and c_j are the weighting and bias parameters, respectively, and x_i is the forecast data normalized by four times the SD of observation σ in the training period from the input layer, as the input range is from -1.0 to 1.0. This output from the jth neuron (y_j) is entered into the output layer as follows:

$$Z = \sum_j^H \left(\tilde{w}_j y_j + \tilde{c} \right) \tag{15}$$

where H is the number of output neurons, and $\tilde{w}_j$ and $\tilde{c}$ are the weighting and bias from the hidden layer to the output layer, respectively. This process is repeated to minimize the cost function, J,

$$J = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(Z_t - O_t \right)^2} \tag{16}$$

where O_t is the observation from time step t, and T is the training period. The final forecast data from ANN using the MLP method are given by

$$\mathrm{MME} = z \times 4\sigma \tag{17}$$

where σ is the observation SD in the training period (Park et $al.$, 2005).

In addition, we use the radial basis function (RBF), which has the same structure as MLP. In the process from the input to the hidden layer of RBF, the jth neuron in the hidden layer is derived by.

$$y_j = \exp\left(-\frac{\sqrt{\left(x - \mu_j\right)^2}}{2\sigma_j^2} \right) \tag{18}$$

where μ_j and σ_j are the mean and SD of each neuron, respectively. The final output of RBF is also calculated as Equation (8) (Lee et $al.$, 2008).

2.2.5 | Genetic algorithm (GA) method

The concept of GA was initially designed by Holland (1975) as a probabilistic technique that explores the optimum solution to a given problem that maximizes or minimizes a particular function by algorithmizing the process of developing and producing the next generation to preserve and survive their genes in an environment of survival of the fittest. GA is applied to MME to find the optimal weight among the models during the training period. Randomly generated initial weights of single models are calculated in the first step of GA by using the hindcast data of each model except the target year. These weights are evaluated by a fitness function to check whether they are optimal or not. If the weight does not provide the optimal solution, the algorithm continues to the next generation through operators such as crossover, mutation, and replacement until an optimal solution is obtained (Henao, 2011). Here, the minimum RMSE between observation and model anomalies is used as the fitness function (Lee et $al.$, 2006; You et $al.$, 2012; Ahn and Lee, 2016). Ahn and Lee (2016) showed that the MME method applying GA to both single-model and MME improved the performance for both winter temperature and precipitation, even over higher-latitude land areas, compared to conventional SCM. This improvement over land seems to be attributable to GA's greater efficiency in finding an optimum solution in a complex region where nonlinear physical properties are evident. Ratnam et $al.$ (2019) also revealed an improvement in seasonal forecast of the 2-m temperature by applying GA over several regions of South America, North America, Australia and Russia compared to the unweighted ensemble mean.

3 | COMPARISON OF MM2E PREDICTION METHODS

We first show the general characteristics and performance of the various deterministic MME methods using the 27 years of hindcasts for 1983–2009. There are four linear (SCM, SLR, MLR, and BSA) and three nonlinear (MLP, RBF, and GA) MME methods.

3.1 | Annual variation of MME prediction methods

Figure 1 shows the annual mean of SDs of the seasonal 2-m temperature in observation and seven MME predictions, together with the observational annual climatology pattern. The annual mean SD of each MME prediction is calculated by averaging the SD of 1-month lead predictions for each season. The temporal variability in observation over the land tends to be much larger than that over the ocean due to the larger thermal inertia of the oceans because of the different surface and atmospheric
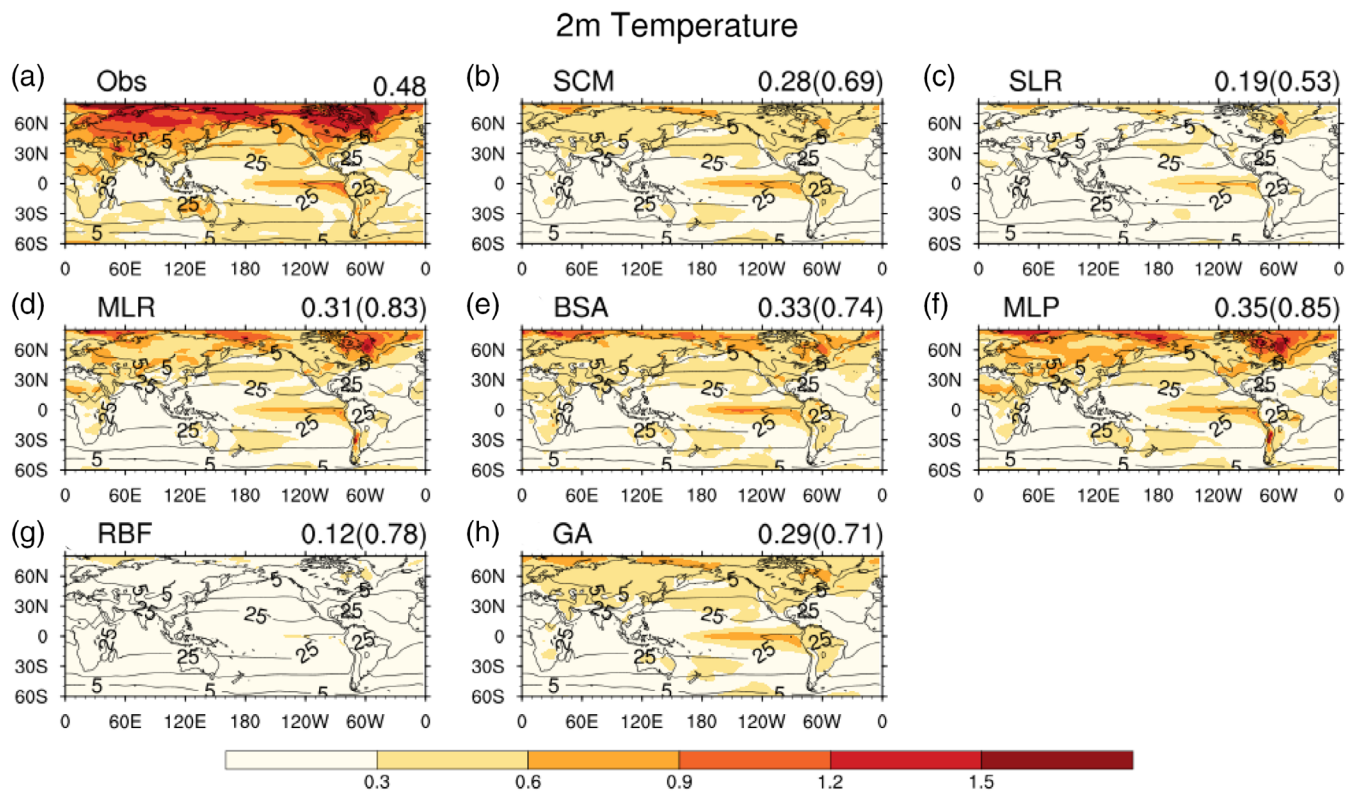
2m Temperature

FIGURE 1 Annual *SD* (shading) of 2-m temperature of (a) observation, (b) SCM, (c) SLR, (d) MLR, (e) BSA, (f) MLP, (g) RBF, and (h) GA for the period of 1983–2009. The contour line indicates the annual climatology of observation for the same period. The unit of climatology for 2-m temperature is °C. The number in the top right corner of each plot shows the mean *SD* over the globe. The pattern correlations between the observation and MME prediction are also given in brackets [Colour figure can be viewed at wileyonlinelibrary.com]

feedbacks that occur over the land compared with the ocean (Stouffer *et al.*, 1994; Bell *et al.*, 2000).

The variability in the model is calculated based on the ensemble mean, and the noise is averaged out; hence, the variability in the ensemble mean is generally smaller than the observed one. Nevertheless, most MME methods represent a contrast of the temporal variability between land and ocean shown in the observation, particularly using the MLP and MLR methods which are highly correlated to the observation (spatial ACCs equal 0.85 and 0.83, correspondingly). There are differences in variability among the MME predictions, most of which are found over land rather than over the ocean because for a 1-month lead, the SST prediction skill for all models is generally high; thus, the differences among the various models are small over the ocean except SLR and RBF. The variability in MME predictions in the equatorial eastern Pacific is relatively higher than that in other regions. Previous studies have shown that most of the interannual variability in the skill of seasonal predictions is affected by oceanic boundary conditions, particularly those associated with the ENSO (Wang *et al.*, 2009; Barnston *et al.*, 2010; Min *et al.*, 2014). The variability associated with oceanic conditions, such as ENSO, is retained in the

MME predictions, leading to a higher variability in the equatorial eastern Pacific.

For precipitation, the largest variability in the observed precipitation spans the tropical region. A similar tendency is found in the MME predictions but with a smaller amplitude (Figure 2) because of ensemble averaging. The results from Figures 1 and 2 reveal that the MME predictions capture the spatial patterns of the observed variability over the ocean quite well, but that the amplitudes of variations from the MME predictions are smaller than those of the observations for both the temperature and precipitation.

3.2 | Interannual and seasonal variation of MME prediction skill

Figures 3 and 4 show the temporal evolution of the forecast skill of MME seasonal predictions for MAM, JJA, SON, and DJF of the 2-m temperature and precipitation, respectively, in terms of ACC during the 27-year period. All of metrics are calculated, including ocean and land. The forecast skill for both variables of MME predictions is related to the ENSO phases (Wang *et al.*, 2009; Barnston *et al.*, 2010; Min *et al.*, 2014).

During the strong 1997/98 El Niño event, the spread of forecast skill among MMEs for precipitation for the decaying phase (MAM 1998) is narrower than that for the maturing phase (DJF 1997), which means that the forecast skills of MMEs are relatively certain after the ENSO events. Previous studies reveal that the atmospheric response associated with the ENSO is lagged by a few months, not only in observation but also in climate models (Trenberth and Paolino, 1981; Trenberth et al., 1998; Kumar and Hoerling, 2003; Bulić and Kucharski, 2012). This may be due to the delayed atmospheric response to the ENSO, in that the oceanic boundary conditions initialized from late January to February (mature phase of El Niño) and used for seasonal prediction for spring, lead to a larger atmospheric response in spring (Kumar and Hoerling, 2003) and a higher forecast skill, regardless of the MME method. The averaged MME forecast performances of both variables during the ENSO events are always better than those during non-ENSO events for all four seasons. Especially, the ACC differences between the ENSO and non-ENSO precipitation events for JJA and DJF are statistically significant at the 5% significance level. The variability in performance among the MME predictions is consistent regardless of the ENSO events (as shown in Figure S13). Although the

ENSO is an important of predictability, the forecast skills of the MME predictions do still vary for both variables. For example, for all seasons GA tends to outperform the other MME methods for both the 2-m temperature and precipitation in terms of not only ACC but also MSSS (Figures 3, 4, and S1–S4). Other oceanic forcing such as El Niño Modoki (Ashok et al., 2007) and Indian Ocean Dipole (Saji et al., 1999) also have an improving effect on MME prediction performance (Figures S14 and S15), but their impacts are lower than the ENSO.

The global averaged TCCs of the MME predictions for the 2-m temperature and precipitation are shown in Figure 5a,b, together with the mean of single model skills (MSMS), which is estimated as the average of the skills of individual models. Generally, the skills of MME predictions for the 2-m temperature are much higher than those for precipitation, and the seasonal skill for the 2-m temperature does not vary greatly. Also, the differences of forecast skills of MME among the seasons are not large and do not show large seasonal dependency. The MSMS for the 2-m temperature is also largely better than that of precipitation. Most MME predictions show higher correlation than the corresponding (based on TCC as a skill metric) MSMS for both the 2-m temperature and precipitation during all four seasons.
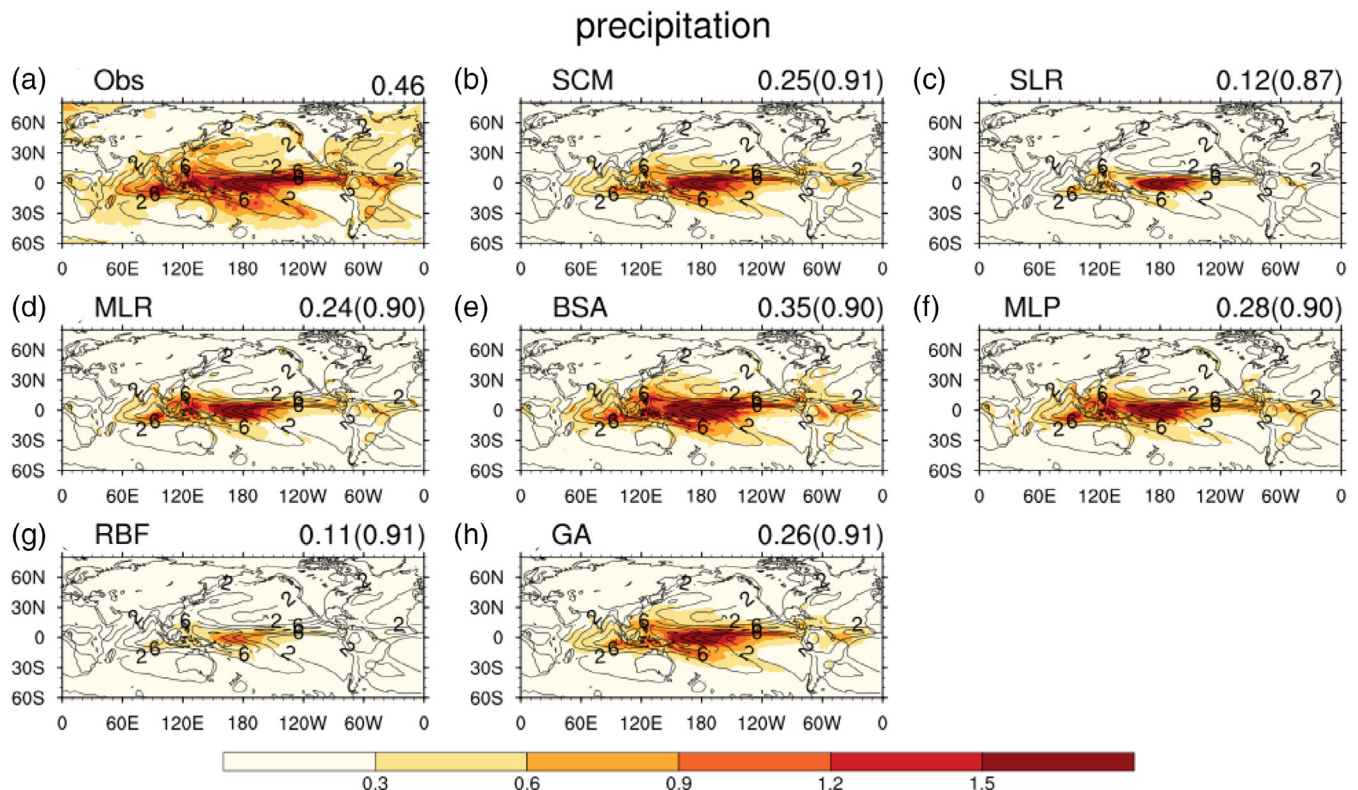

FIGURE 2 Same as Figure 1 but for precipitation. The unit of climatology for precipitation is mm·day^{-1} [Colour figure can be viewed at wileyonlinelibrary.com]
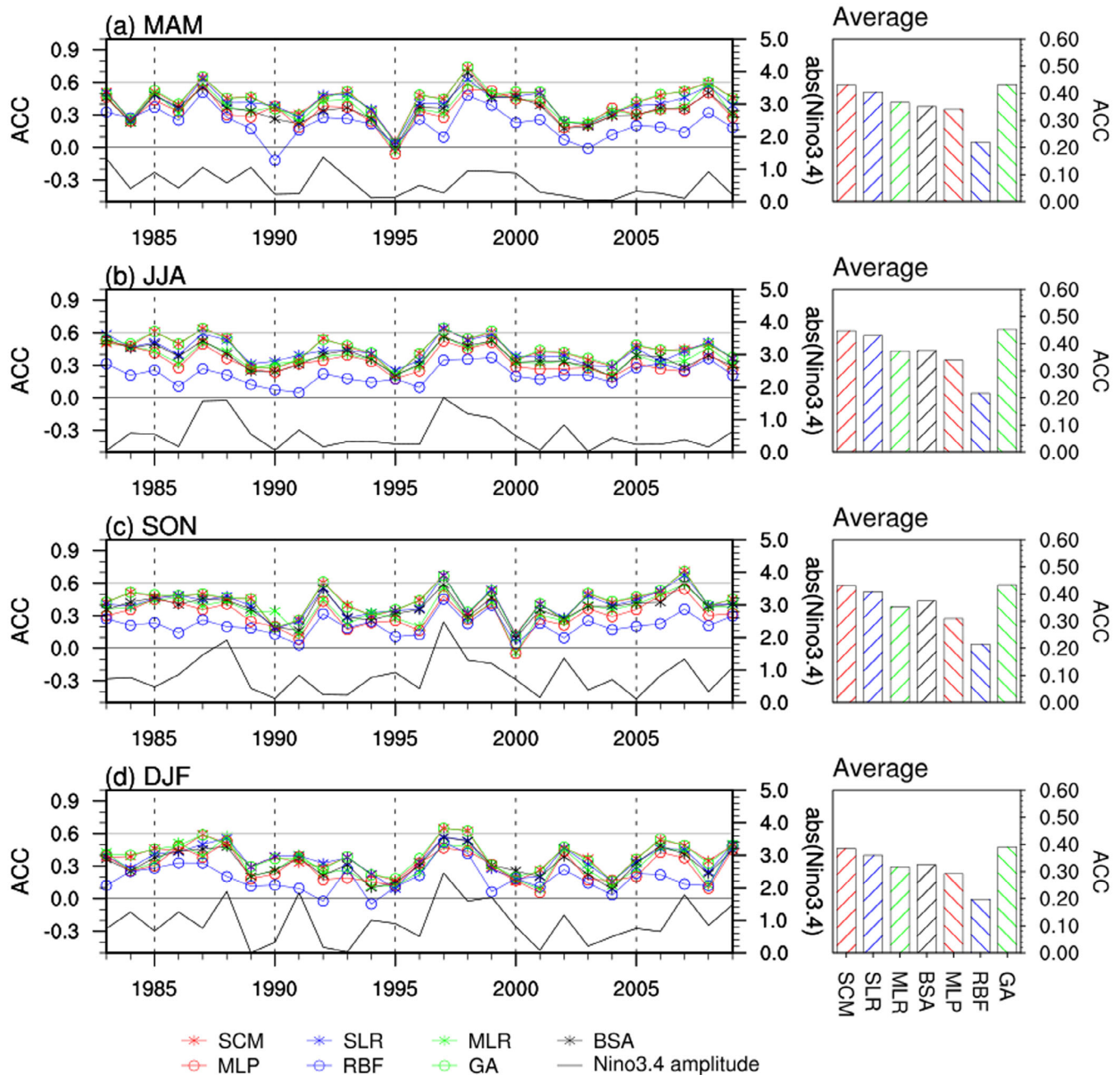
FIGURE 3 Time series of ACCs between observations and predictions of 2-m temperature by each MME method over the globe (coloured lines with marker) and absolute value of Niño 3.4 index (black line) for (a) MAM, (b) JJA, (c) SON, and (d) DJF for the period 1983–2009. The average scores during the same period are also shown [Colour figure can be viewed at wileyonlinelibrary.com]

However, the MLP method reveals lower correlations than MSMS for both variables during all four seasons, indicating that not all MME methods improve the skill compared with MSMS.

Figure 6a,b show the global averaged RMSE for the 2-m temperature and precipitation, respectively. Generally, the forecast errors of MME are larger in boreal winter than in boreal summer, because of higher variability in winter in extratropical Northern Hemisphere land regions, particularly for the 2-m temperature. The

forecast errors averaged over separate single model forecasts for the 2-m temperature are 0.81, 0.65, 0.79, and 1.01 during the MAM, JJA, SON, and DJF seasons, respectively. Most MME predictions show lower RMSEs than the corresponding (based on RMSE as a skill metric) MSMSs do, except for MLP during MAM, SON, and DJF. Similar to the 2-m temperature, the RMSEs of all MME predictions for precipitation are smaller than the corresponding MSMS, that is, the MME prediction tends to reduce the forecast errors.
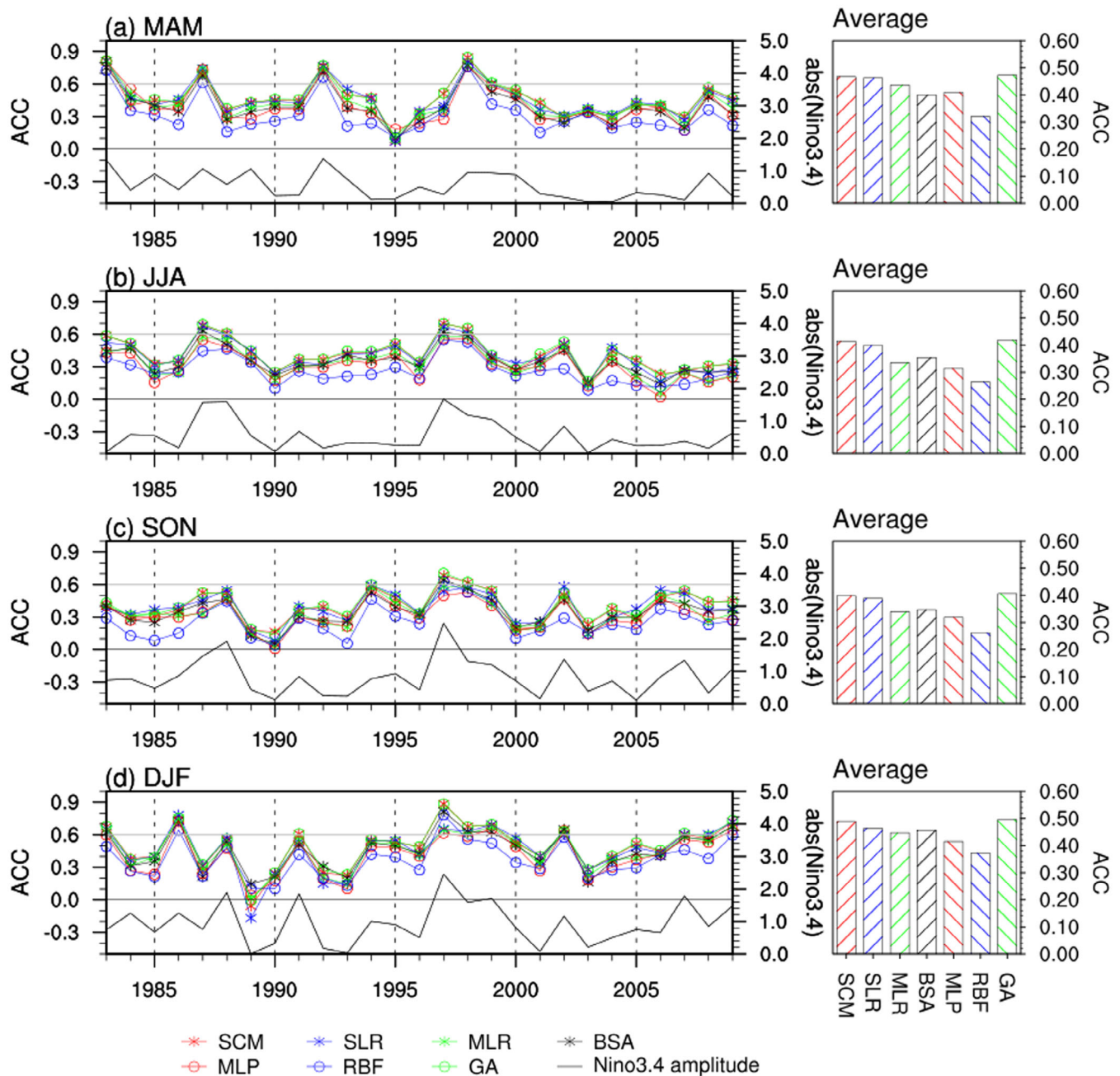
FIGURE 4 Same as Figure 3 but for precipitation [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | Overall performance of the MME prediction in the retrospective period

As shown in the Taylor diagram (Taylor, 2005) in Figure S16, the MME predictions mostly underestimate the interannual variation of both the 2-m temperature and precipitation for all four seasons worldwide. The TCC and normalized *SD* for each MME prediction show very low seasonal variation. Figure 7 is presented to comprehensively determine the 4-season-averaged TCC and RMSE of each MME and individual model for the 2-m

temperature and precipitation. First, for the 2-m temperature, the skill and RMSE of the MME predictions generally have a linear relationship, as opposed to that of single models. Second, the performance of MMEs is superior to that of the individual models (e.g., Min *et al.*, 2014; Kim *et al.*, 2016). GA has good forecast performance, similar to SCM, among all MMEs and individual models with the highest TCC (0.51), which is significant at the 99% confidence level, and the lowest RMSE (0.74). The worst is MLP, which has a TCC of 0.41 and RMSE of 0.80.
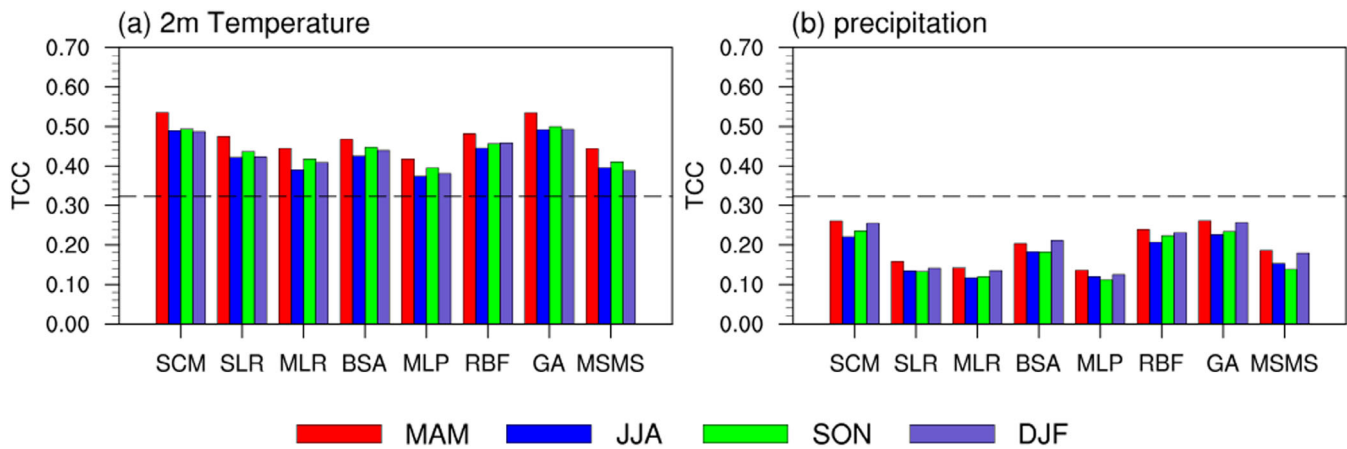
FIGURE 5 Global averaged temporal correlations between observations and predictions of (a) 2-m temperature and (b) precipitation by each MME method and single-model's mean (SMM) for MAM, JJA, SON, and DJF for the period 1983–2009. Dashed lines indicate that the estimated score is statistically significant at the 5% level using the one-tailed Student's t test [Colour figure can be viewed at wileyonlinelibrary.com]
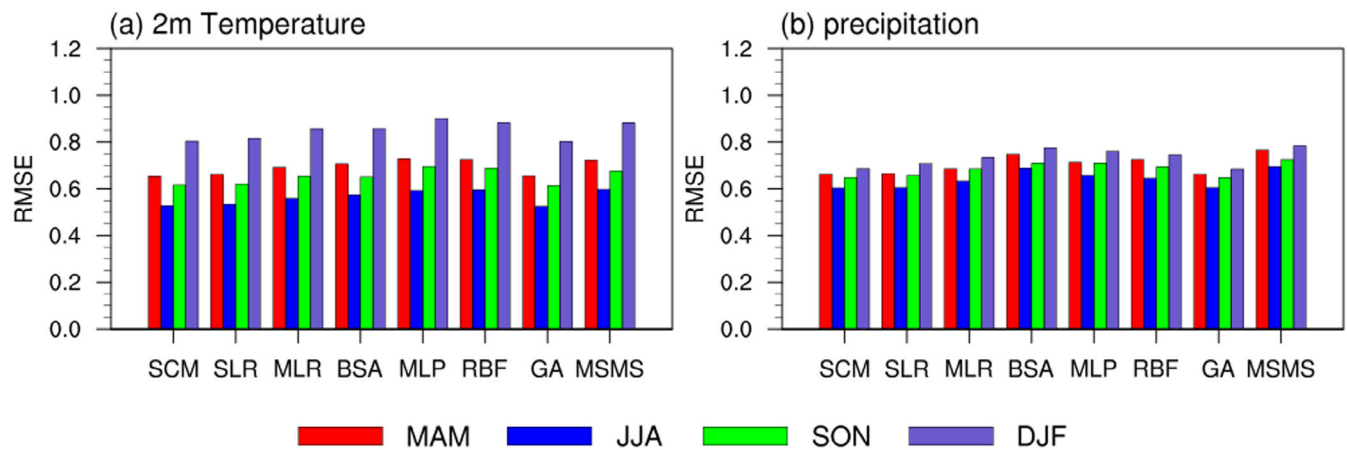

FIGURE 6 Global averaged RMSE of (a) 2-m temperature and (b) precipitation predictions of each MME method and single-model's mean (SMM) for MAM, JJA, SON, and DJF for the same period [Colour figure can be viewed at wileyonlinelibrary.com]

For precipitation, the TCC and RMSE do not show a distinct linear relationship, but there is a clear difference in performance among the MMEs and single models. GA also exhibits good performance for precipitation with TCC of 0.20 and RMSE of 0.52. MLP is also ranked as the worst MME prediction for precipitation with poor forecast skill (0.12) and large forecast error (0.55). The forecast performance of GA is consistently comparable to that of SCM, which is widely used in many operational centres for each variable and each season.

The mean square skill score (MSSS) which is recommended by the WMO Standardize Verification System (SVS) for LRFs can be decomposed as phase errors (through the correlation between forecast and observation), amplitude errors (through the ratio of the forecast to observed SDs) and overall bias error, respectively. If the model or MME performance is good with a strong

correlation and small amplitude error, the forecast can be considered suitable for use in forecasts. A good forecast is achieved when the ratio of the forecasted to observed SDs is close to one, overall bias through the difference between the forecast and observation is near zero, and the MSSS is close to one. Negative or zero MSSS values indicate that deterministic forecasts are worse than or the same as climatology forecast (WMO, 2010, 2018; Setiawan et al., 2017). In terms of MSSS, GA outperforms the other MMEs for the 2-m temperature in both boreal summer and winter (as shown in Figures S1 and S2). The correlation does not vary significantly among all MMEs (Figures S5 and S6). However, the ratio of SDs for RBF is close to zero worldwide (Figures S9 and S10). As a result, the MSSS values of RBF are lower than those of the other methods. For the precipitation, the skilful MSSS mainly shows in the Maritime continent and equatorial Pacific
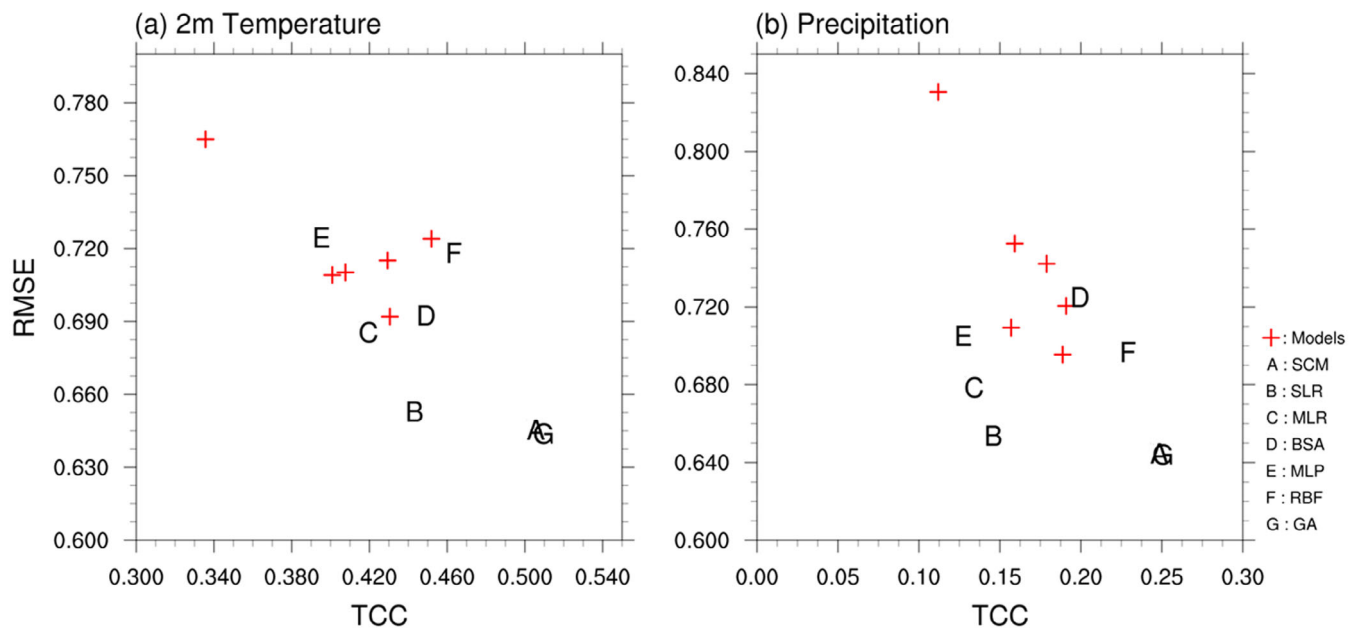
FIGURE 7 Diagram of temporal correlation-RMSE of each MME method of (a) 2-m temperature and (b) precipitation averaged over four seasons (MAM, JJA, SON, and DJF). The red crosses indicate each individual model [Colour figure can be viewed at wileyonlinelibrary.com]

for both JJA and DJF seasons (Figures S3 and S4) for all MMEs and it is related to the correlation. The large ratio of the *SD*s causes MSSS to be negative, as clearly demonstrated by BSA of the precipitation for the JJA and DJF (Figures S11 and S12).

4 | PERFORMANCE OF THE GA METHOD OVER RCOF REGIONS

As discussed above, the prediction skill of GA is almost equal to that of SCM for the hindcast period worldwide. As the performance of SCM over the RCOF regions was already revealed in a previous study (Kim *et al.*, 2016), we investigate the forecast performance of the GA method over the RCOF regions in this section.

RCOF regions have been designated by the WMO and the corresponding domains are listed in Table 2. RCOF produces consensus-based, user-relevant climate outlook products in real-time to reduce climate-related risks and support sustainable development for the region in question (WMO, 2016). One of the primary sources of information for the climate outlook of RCOF is the real-time seasonal forecast produced by WMO LC-LRFMME.

Figure 8 shows the scatter diagram between TCC and normalized RMSE (NRMSE) of the most reliable MME prediction for the hindcast period, GA, for the 2-m temperature and precipitation for each season, averaged over the domain of each RCOF region. Here, we calculate the NRMSE, which is obtained from the RMSE divided by

the *SD* of the observation, to compare the forecast errors of the 2-m temperature and precipitation together. The seasonality of the forecast performance over most RCOF regions is not large for both 2-m temperature and precipitation. The most consistent result is that the forecast skills and RMSE for the 2-m temperature are better than those for precipitation, showing a larger TCC. In addition, the forecast performance still varies based on region or variable, even though the most reasonable MME prediction method (i.e., GA) is used. In terms of MSSS as well, GA performs better than the other methods in most RCOF regions, especially for boreal winter 2 m temperature and precipitation (Figure S21). The forecast skills of both the variables decrease toward the extratropical RCOF regions. Previous studies have revealed that the prediction skill for individual models is greater over the tropics than the extratropics and greater over ocean than land (Wang *et al.*, 2009; Peng *et al.*, 2011; Kim *et al.*, 2012; Lee *et al.*, 2014; Ham *et al.*, 2019). GA also outperforms in tropical RCOF regions, such as ASEANCOF, PICOF, and PRESAO, which are more affected by the interannual variations under oceanic boundary conditions, for example, ENSO (e.g., Ropelewski and Halpert, 1989; Mason and Goddard, 2001; Jin *et al.*, 2008). However, the skills over extratropical RCOF regions, not affected or weakly affected by the ENSO, remain relatively low (Alexander *et al.*, 2002; Palmer *et al.*, 2004; Kryjov, 2012; Kumar *et al.*, 2013; Min *et al.*, 2014). This conforms to results from many previous studies that the skillful forecasts are

TABLE 2 Domains of the RCOF regions (see figure 10 in Kim *et al.*, 2016)

RCOF	Domain
GHACOF (Greater Horn of Africa Climate Outlook Forum)	22.5–50°E, 12.5S–22.5°N
SARCOF (Southern African Regional Climate Outlook Forum)	10–50°E, 35°S–7.5°N
PRESAO (Climate Outlook Forum for West Africa)	17.5°W–27.5°E, 0–27.5°N
FOCRAII (Forum on Regional Climate Monitoring, Assessment and Prediction for Regional Association II)	50–150°E, 10°S–50°N
WCSACOF (Western Coast of South America Climate Outlook Forum)	80–57.5°W, 55°S–12.5°N
SSACOF (Southeast of South America Climate Outlook Forum)	70–40°W, 40–20°S
PICOF (Pacific Islands online Climate Outlook Forum)	140°E–25°W, 25°S–5°N
CACOF (Climate Outlook Forum for Central America)	92.5–77.5°W, 5–17.5°N
SEECOF (Southeastern Europe Climate Outlook Forum)	10–40°E, 30–50°N
SASCOF (South Asian Climate Outlook Forum)	60–100°E, 0–40°N
CARICOF (Caribbean Climate Outlook Forum)	90–60°W, 10–30°N
NEACOF (North Eurasian Climate Outlook Forum)	20–180°E, 40–70°N
ASEANCOF (ASEAN Climate Outlook Forum)	90–140°E, 10°S–30°N

concentrated in the seasons and regions affected by the ENSO (e.g., Ropelewski and Halpert, 1989; Mason and Goddard, 2001; Jin *et al.*, 2008) as well as in the extratropical regions strongly related to the ENSO via teleconnections during its peak phases (Wang *et al.*, 2009; Barnston *et al.*, 2010).

5 | REAL-TIME FORECAST

Next, we investigate the seasonal predictability of different MME systems for the 1-month lead seasonal 2-m temperature and precipitation, using real-time forecasts for the period during MAM 2012-DJF 2015/16. Figure 9 shows the real-time forecast skills of the two MME predictions (GA and SCM) for the real-time forecast period and the spread of the skill of all MMEs for the 2-m temperature and precipitation, in terms of ACCs. We focus solely on these two MME methods because we previously assessed GA as the most skilful MME method for the hindcast period, and SCM is widely used in many operational centres. Overall, the real-time forecast skill for both the 2-m temperature and precipitation increased in recent years, as it was affected by the very strong El Niño in 2015. As mentioned above for the hindcast period, the real-time forecast skills of all MME methods are also highly related to ENSO variability. The forecast skill among all seven MMEs ranges from 0.35 to 0.63 for the 2-m temperature and from 0.77 to 0.82 for the precipitation in DJF 2015. The forecast skills for GA and SCM are not always better than those of the other MME predictions for the period of real-time forecast; however, they consistently show relatively higher performance, regardless of the season. The time-averaged ACCs of GA and SCM are 0.43 and 0.42 (0.31 and 0.32) for the 2-m temperature (precipitation). There are few skill differences between GA and SCM in the real-time forecast period, although GA consistently slightly outperforms other MME methods for the hindcast period.

The forecast errors, in terms of RMSEs, of the 2-m temperature for the same period are illustrated in Figure 10a. The most notable feature is that the skill differences among the MME methods during DJF are larger than those during the other seasons. The RMSE of the MLP shows the greatest seasonal variation, with a mean and *SD* of 1.44 and 0.56, respectively. The increased forecast error during DJF may be due to the wider error spread of individual models compared with that in the other seasons (figure not shown). In general, the forecast error of both GA and SCM is lower (with time-averaged RMSE of 0.94 for GA and 0.95 for SCM) than that of other MMEs, even in winter. Figure 10b shows the forecast error for precipitation for the real-time forecast period. Among the MME methods, the time-averaged forecast error ranges from 0.80 to 1.09. The forecast errors of GA and SCM with time-averaged RMSE of 0.88 and 0.85, respectively, are relatively low among those of the MME methods. However, SLR shows the lowest error with a time-averaged RMSE of 0.80.

These two MME systems present lower errors than those of the other MMEs in most cases, when the error spreads of MMEs are quite large (e.g., precipitation for SON 2013). The errors of the two MME predictions are not largely different in the real-time forecast; the time averaged RMSEs of GA and SCM are both 0.94 for the 2-m temperature and 0.88 and 0.85 for precipitation, respectively. Individual models show better performance during the very strong El Niño event in 2015 than in other years.
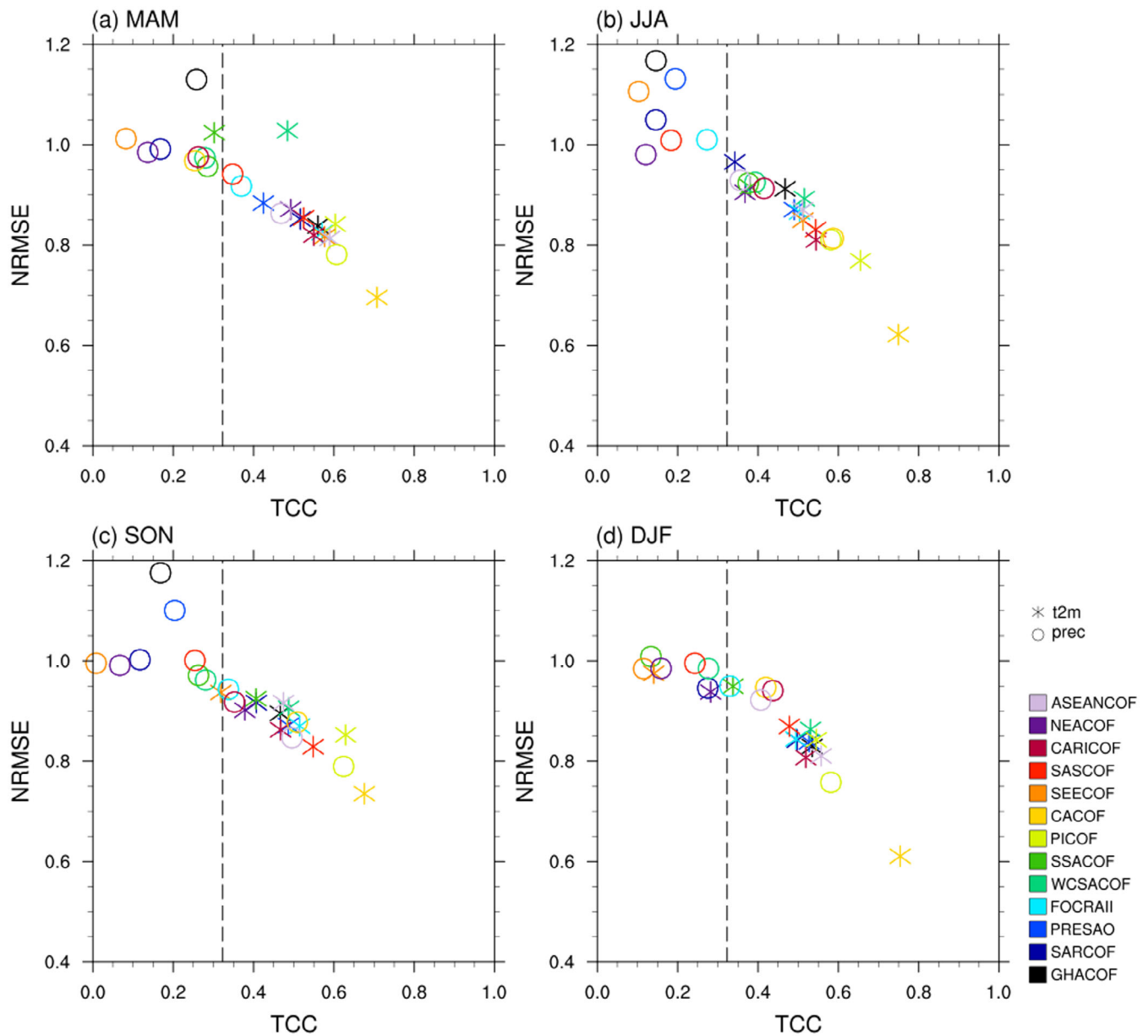
FIGURE 8 Diagram of TCC-NRMSE of GA, which is selected as the most skilful MME method for the RCOF regions during (a) MAM, (b) JJA, (c) SON, and (d) DJF mean 2-m temperature and precipitation for the period of 1983–2009. Dashed line indicates that the estimated TCC is statistically significant at the 5% level using the one-tailed Student's *t* test [Colour figure can be viewed at wileyonlinelibrary.com]

Nonetheless, both SCM and GA methods outperform all the individual models. The strong El Niño events (1997/98 El Niño in the hindcast period and 2015 El Niño in the real-time forecast period) certainly improve prediction performance from the autumn of the El Niño year to the spring of the following year (Figures S22 and S23).

6 | DISCUSSION AND SUMMARY

The MME methods have been developed and evaluated individually, sporadically and in isolation, so it is

necessary to assess their overall performance comprehensively and coherently. We evaluated the global skills of MMEs for the hindcast period and suggested the most reasonable MME method. For the assessed variables, most MME predictions present reasonable skills in terms of spatial patterns of annual variation and temporal evolution for all four seasons for the 27-year hindcast periods, particularly during the El Niño events. The ENSO variability was one of the main sources of predictability, not only in the individual models but also in the MME prediction (Lee *et al.*, 2011, 2013a, 2013b; Min *et al.*, 2014, 2017). The forecast skills and errors, in terms

FIGURE 9 Time series of ACC between the observations and predictions for 1-month lead seasonal mean (a) 2-m temperature and (b) precipitation for GA (red) and SCM (blue) for the real-time period MAM 2012–DJF 2015/16 over the globe. The grey shaded area shows the spread of scores for all MME methods for the same period [Colour figure can be viewed at wileyonlinelibrary.com]
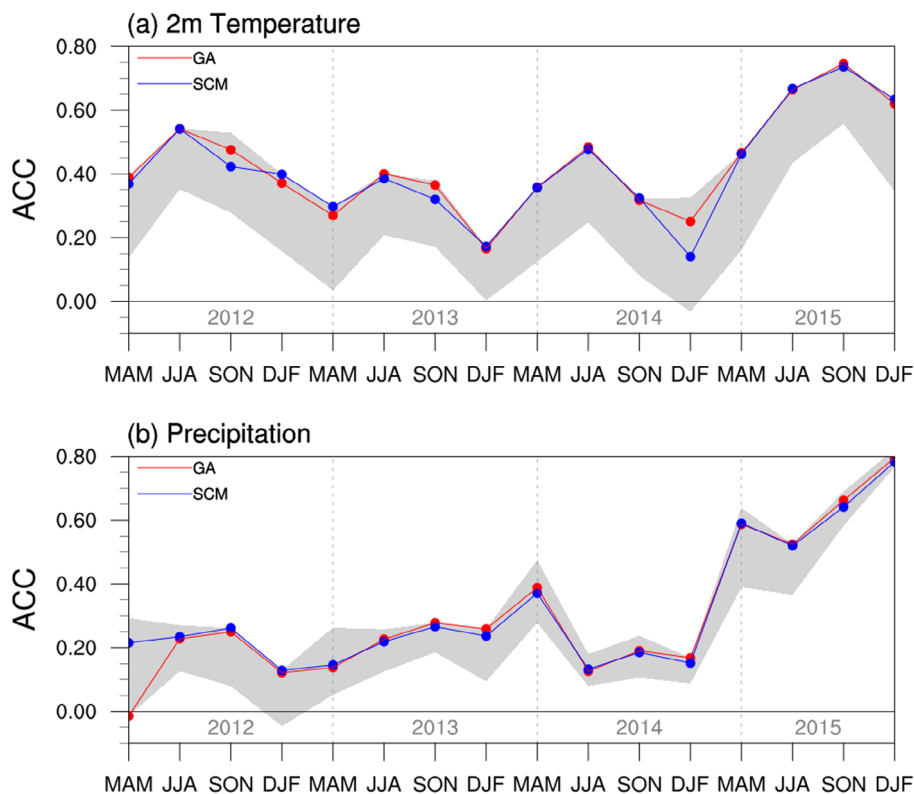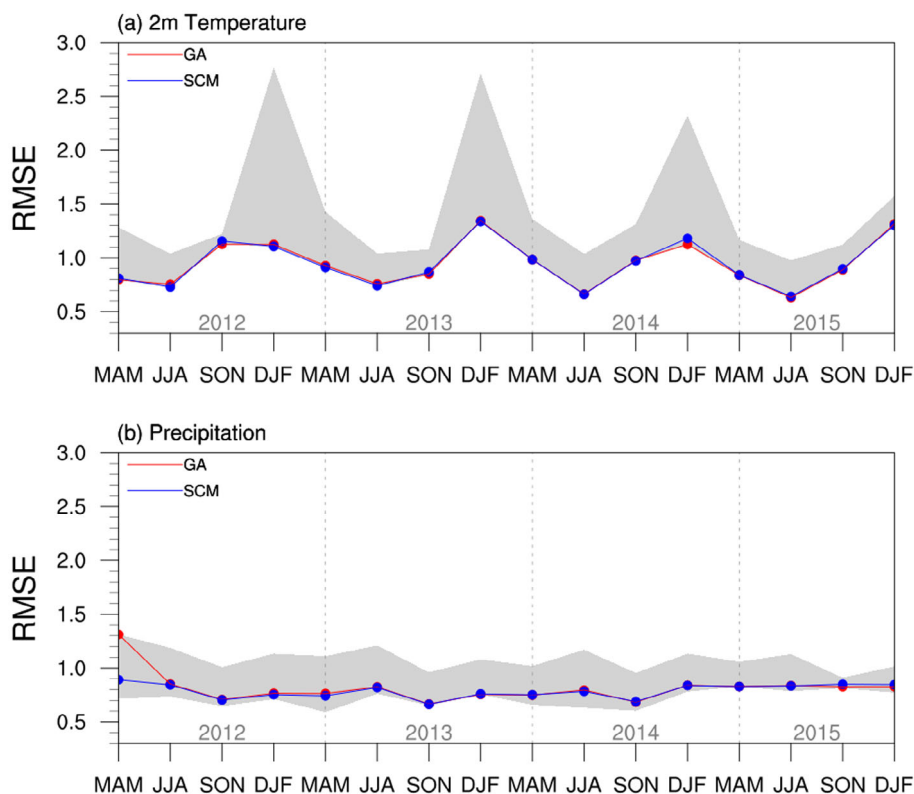

FIGURE 10 Time series of RMSE between the observations and predictions for 1-month lead seasonal mean (a) 2-m temperature and (b) precipitation for GA (red) and SCM (blue) for the same real-time period over the globe. The grey shaded area shows the spread of scores for all MME methods for the same period [Colour figure can be viewed at wileyonlinelibrary.com]

of global TCC and RMSE, of some MME methods were poor compared with those of MSMS, and we take GA as the most reasonable MME prediction method based on

performance. Even with the most reasonable MME forecasting method (i.e., GA), the forecast performance still depends on the region or variable. In the real-time

forecast period, the two MME systems (GA and SCM) reveal consistently higher skills and lower errors compared to the other methods for the global 2-m temperature and precipitation.

The following study caveats should be noted. First, the forecast skills vary depending on the region or variable, even in the most skilful MME system. Also, the forecast skill for precipitation remains poorer than that for the 2-m temperature, even using various MME methods. This is due to the complex physics of local-scale precipitation, resulting in more noise and therefore unpredictable characteristics (Gong *et al.*, 2003; Min *et al.*, 2017). In addition, the forecast errors are quite large in the extratropics for both the 2-m temperature and precipitation. Second, the MME predictions outperform those by single models in both hindcast and real-time forecast. The differences of forecast skills between GA and SCM were not statistically significant, but GA remains the most skilful MME prediction for the 2-m temperature in the real-time forecast.

The limited seasonal prediction skills of the MME methods can be partially explained using the Empirical Orthogonal Function (EOF) analysis. The first mode of EOF of observation for the global 2-m temperature is close to the global warming pattern, and the first principal component (PC1) is highly correlated (0.66) with the global warming signal. This mode explains about 15.8% of the total variance for 1983–2009 (Figures S17a and S18a). The second EOF mode of observation for the global 2-m temperature, which explains 13.8% of the variance, mainly shows the Arctic Oscillation (AO)-related impact over the land in the Northern Hemisphere (Kryzhov and Gorelits, 2015; He *et al.*, 2017). The correlation coefficient between PC2 and the AO index obtained from CPC (available at http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao_index.html) is 0.52 (Figures S19a and S20a). The first EOF modes of most MME methods except for RBF are related to global warming. The second EOF mode of the MMEs (except for RBF) is associated with the ENSO rather than with the AO (Figures S19b–h and S20b–h).

Several studies have attributed the difficulty in assessing MME methods to the small samples available to calculate optimal weights (Kharin and Zwiers, 2002; Kryjov *et al.*, 2006; DelSole and Shukla, 2009; Rodrigues *et al.*, 2014; Min *et al.*, 2017) and to overfitting during the training period (Davis, 1978; Michaelsen, 1987; Min *et al.*, 2014). Nevertheless, it is relevant that we explore the possibility of finding MME techniques other than SCM for use in operational seasonal prediction. For RCOF regions, it is also important to utilize an objective MME approach instead of a consensus-based subjective approach for developing seasonal forecasts. GA shows good performance in the RCOF regions, especially in the tropical RCOF regions, such as ASEANCOF, PICOF, and PRESAO, which are directly affected by oceanic boundary conditions (e.g., ENSO). As noted in the introduction, the study results will be helpful to improve the seasonal outlook for RCOF regions, in terms of both deterministic and probabilistic predictions. Improved deterministic forecast can be used as the mean of forecast PDF, and the variance of forecast PDF can be assessed by accounting for both ensemble spread and calibration error as suggested by Min *et al.* (2011) and Kryzhov (2012).

ORCID

Joong-Bae Ahn 🄳 https://orcid.org/0000-0001-6958-2801

REFERENCES

Adler, R.F., Huffman, G.J., Chang, A., Ferraro, R., Xie, P.P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P. and Nelkin, E. (2003) The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology*, 4(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Ahn, J.-B. and Lee, J. (2016) A new multimodel ensemble method using nonlinear genetic algorithm: An application to boreal winter surface air temperature and precipitation prediction. *Journal of Geophysical Research: Atmospheres*, 121(16), 9263–9277. https://doi.org/10.1002/2016JD025151.

Ahn, J.-B., Lee, J. and Im, E.-S. (2012) The reproducibility of surface air temperature over South Korea using dynamical downscaling and statistical correction. *Journal of the Meteorological Society of Japan*, 90(4), 493–507. https://doi.org/10.2151/jmsj.2012-404.

Alexander, M.A., Bladé, I., Newman, M., Lanzante, J.R., Lau, N.C. and Scott, J.D. (2002) The atmospheric bridge: the influence of ENSO teleconnections on air-sea interaction over the global oceans. *Journal of Climate*, 15(16), 2205–2231 https://doi.org/10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

Ashok, K., Behera, S.K., Rao, S.A., Weng, H. and Yamagata, T. (2007) El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, 112(11), 1–27. https://doi.org/10.1029/2006JC003798.

Barnston, A.G., Li, S., Mason, S.J., Dewitt, D.G., Goddard, L. and Gong, X. (2010) Verification of the first 11 years of IRI's seasonal climate forecasts. *Journal of Applied Meteorology and Climatology*, 49(3), 493–520. https://doi.org/10.1175/2009JAMC2325.1.

Bell, J., Duffy, P., Covey, C., Sloan, L. and Investigators, C. (2000) Comparison of temperature variability in observations and

sixteen climate model simulations. *Deleted DOIs*, 27(2), 261–264. https://doi.org/10.1029/1999gl008060.

Bulić, I.H. and Kucharski, F. (2012) Delayed ENSO impact on spring precipitation over North/Atlantic European region. *Climate Dynamics*, 38(11–12), 2593–2612. https://doi.org/10.1007/s00382-011-1151-9.

Davis, R.E. (1978) Predictability of sea level pressure anomalies over the North Pacific Ocean. *Journal of Physical Oceanography*, 8, 233–246.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N. and Vitart, F. (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. https://doi.org/10.1002/qj.828.

DelSole, T. (2007) A Bayesian framework for multimodel regression. *Journal of Climate*, 20(12), 2810–2826. https://doi.org/10.1175/JCLI4179.1.

DelSole, T. and Shukla, J. (2009) Artificial skill due to predictor screening. *Journal of Climate*, 22(2), 331–345. https://doi.org/10.1175/2008JCLI2414.1.

Gong, X, Barnston, AG, Ward, MN. 2003. The effect of spatial aggregation on the skill of seasonal precipitation forecasts. *Journal of Climate*, 16(18): 3059–3071. https://doi.org/10.1175/1520-0442(2003)016<3059:TEOSAO>2.0.CO;2.

Graham, R.J., Yun, W.-T., Kim, J., Kumar, A., Jones, D., Bettio, L., Gagnon, N., Kolli, R.K. and Smith, D. (2011) Long-range forecasting and the global framework for climate services. *Climate Research*, 47(1), 47–55. https://doi.org/10.3354/cr00963.

Hagedorn, R., Doblas-Reyes, F.J. and Palmer, T.N. (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 57(3), 219–233. https://doi.org/10.1111/j.1600-0870.2005.00103.x.

Ham, S., Lim, A.-Y., Kang, S., Jeong, H., Jeong, Y., Wang, B., Xiang, B., Wu, S. and Fu, J.X. (2019) A newly developed APCC SCoPS and its prediction of East Asia seasonal climate variability. *Climate Dynamics*, 53(5), 3703–3704. https://doi.org/10.1007/s00382-019-04894-y.

He, S., Gao, Y., Li, F., Wang, H. and He, Y. (2017) Impact of Arctic oscillation on the east Asian climate: a review. *Earth-Science Reviews*, 164, 48–62. https://doi.org/10.1016/j.earscirev.2016.10.014.

Henao, J.D.V. (2011) An introduction to chaos-based algorithms for numerical optimization. *Avances en Sistemas e Informática*, 8(1), 51–60.

Holland, J.H. (1975) Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. In: *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Oxford, England: U Michigan Press.

Janowiak, J.E. and Xie, P. (1999) CAMS-OPI: a global satellite-rain gauge merged product for real-time precipitation monitoring applications. *Journal of Climate*, 12, 3335–3342 https://doi.org/10.1175/1520-0442(1999)012<3335:COAGSR>2.0.CO;2.

Jin, E.K., Kinter, J.L., Wang, B., Park, C.K., Kang, I.S., Kirtman, B. P., Kug, J.S., Kumar, A., Luo, J.J., Schemm, J., Shukla, J. and Yamagata, T. (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Climate Dynamics*, 31(6), 647–664. https://doi.org/10.1007/s00382-008-0397-3.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J.E., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R. W., Jenne, R. and Joseph, D. (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77, 437–471.

Kang, I.-S., Lee, J.-Y. and Park, C.-K. (2004) Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. *Journal of Climate*, 17(4), 834–844 https://doi.org/10.1175/1520-0442(2004)017<0834:PPOSMP>2.0.CO;2.

Ke, Z., Dong, W., Zhang, P. and Zhao, T. (2009) An analysis of the difference between the multiple linear regression approach and the multimodel ensemble mean. *Advances in Atmospheric Sciences*, 26, 1157–1168.

Kharin, V.V. and Zwiers, F.W. (2002) Climate predictions with multimodel ensembles. *Journal of Climate*, 15(7), 793–799 https://doi.org/10.1175/1520-0442(2002)015<0793:CPWME>2.0.CO;2.

Kim, G., Ahn, J.-B., Kryjov, V.N., Sohn, S.-J., Yun, W.-T., Graham, R., Kolli, R.K., Kumar, A. and Ceron, J.-P. (2016) Global and regional skill of the seasonal predictions by WMO Lead Centre for Long-Range Forecast Multi-Model Ensemble. *International Journal of Climatology*, 36(4), 1657–1675. https://doi.org/10.1002/joc.4449.

Kim, H.-M., Webster, P.J. and Curry, J.A. (2012) Seasonal prediction skill of ECMWF system 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. *Climate Dynamics*, 39(12), 2957–2973. https://doi.org/10.1007/s00382-012-1364-6.

Kim, O.-Y. (2018) Assessment of seasonal prediction of South Pacific Convergence Zone using APCC multi-model ensembles. *Climate Dynamics*, 50(9–10), 3237–3250. https://doi.org/10.1007/s00382-017-3802-y.

Kim, O.-Y. and Chan, J.C.L. (2018) Cyclone-track based seasonal prediction for South Pacific tropical cyclone activity using APCC multi-model ensemble prediction. *Climate Dynamics*, 51(9–10), 3209–3229. https://doi.org/10.1007/s00382-018-4075-9.

Kim, S.T., Sohn, S.-J. and Kug, J.-S. (2017) Winter Temperatures over the Korean Peninsula and East Asia: Development of a New Index and its Application to Seasonal Forecast. *Climate Dynamics*, 49(5-6), 1567–1581. https://doi.org/10.1007/s00382-016-3402-2.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L., Paolino, D.A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M.P., Becker, E., Peng, P., Tripp, P., Huang, J., Dewitt, D.G., Tippett, M.K., Barnston, A.G., Li, S., Rosati, A., Schubert, S.D., Rienecker, M., Suarez, M., Li, Z.E., Marshak, J., Lim, Y.K.,

Tribbia, J., Pegion, K., Merryfield, W.J., Denis, B., Wood, E.F. (2014). The north American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(April): 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1.

KMA. (2009a) *Development of Advanced Climate Prediction Technique (Multi-Model Ensemble) (in Korean)*. Korea: Seoul.

KMA. (2009b) *Intra-Seasonal Variability Analysis Using Asia-Pacific Climate Model Ensemble and Development of Application Technique (in Korean)*. Korea: Seoul.

Krishnamurti, T.N., Kishtawal, C.M., LaRow, T.E., Bachiochi, D.R., Zhang, Z., Williford, C.E., Gadgil, S. and Surendran, S. (1999) Improved weather and seasonal climate forecasts from multi-model superensemble. *Science*, 285(1548), 1–4. https://doi.org/10.1126/science.285.5433.1548.

Krishnamurti, TN, Kishtawal, CM, Zhang, Z, LaRow, T, Bachiochi, D, Williford, E, Gadgil, S, Surendran, S. 2000. Multi-model ensemble forecasts for weather and seasonal climate. *Journal of Climate*, 13(23): 4196–4216. https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2.

Kryjov, V.N. (2012) Seasonal climate prediction for North Eurasia. *Environmental Research Letters*, 7(1), 1–6. https://doi.org/10.1088/1748-9326/7/1/015203.

Kryjov, V.N., Yang, H., Nohara, D., Song, B.-G., Lee, D.-Y., An, K.-H., Shon, S.-J., Min, Y.-M., Hameed, S.N. and Tam, C.-Y. (2006) Assessment of the climate forecasts produced by individual models and MME methods. In: *APCC Technical Report*. Busan: South Korea.

Kryzhov, V.N. (2012) Downscaling of the global seasonal forecasts of hydrometcenter of Russia for North Eurasia. *Russian Meteorology and Hydrology*, 37(5), 291–297. https://doi.org/10.3103/S1068373912050019.

Kryzhov, V.N. and Gorelits, O.V. (2015) The arctic oscillation and its impact on temperature and precipitation in Northern Eurasia in the 20th century. *Russian Meteorology and Hydrology*, 40(11), 711–721. https://doi.org/10.3103/S1068373915110011.

Kumar, A., Chen, M. and Wang, W. (2013) Understanding prediction skill of seasonal mean precipitation over the tropics. *Journal of Climate*, 26(15), 5674–5681. https://doi.org/10.1175/JCLI-D-12-00731.1.

Kumar, A. and Hoerling, M.P. (2003) The nature and causes for the delayed atmospheric response to El Niño. *Journal of Climate*, 16(9), 1391–1403. https://doi.org/10.1175/1520-0442-16.9.1391.

Lee, D.Y., Ahn, J.-B. and Ashok, K. (2013a) Improvement of multi-model ensemble seasonal prediction skills over east Asian summer monsoon region using a climate filter concept. *Journal of Applied Meteorology and Climatology*, 52(5), 1127–1138. https://doi.org/10.1175/JAMC-D-12-0123.1.

Lee, D.Y., Ahn, J.-B., Ashok, K. and Alessandri, A. (2013b) Improvement of grand multi-model ensemble prediction skills for the coupled models of APCC/ENSEMBLES using a climate filter. *Atmospheric Science Letters*, 14(3), 139–145. https://doi.org/10.1002/asl2.430.

Lee, D.Y., Ashok, K. and Ahn, J.-B. (2011) Toward enhancement of prediction skills of multimodel ensemble seasonal prediction: a climate filter concept. *Journal of Geophysical Research*, 116(D6), D06116. https://doi.org/10.1029/2010JD014610.

Lee, G.-H., Choi, H. and Jung, S. (2008) Control of an angle and a position of inverted pendulum system using TS-fuzzy compensated by RBF neural network (in Korean). *KIIS Fall Conference*, 18(2), 297–300.

Lee, M.-I., Kang, H.-S., Kim, D., Kim, D., Kim, H. and Kang, D. (2014) Validation of the experimental hindcasts produced by the GloSea4 seasonal prediction system. *Asia-Pacific Journal of Atmospheric Sciences*, 50(3), 307–326. https://doi.org/10.1007/s13143-014-0019-4.

Lee, Y.H., Park, S.K. and Chang, D.-E. (2006) Parameter estimation using the genetic algorithm and its impact on quantitative precipitation forecast. *Annales Geophysicae*, 24(12), 3185–3189. https://doi.org/10.5194/angeo-24-3185-2006.

Marius-Constantin, P., Balas, V.E., Perescu-Popescu, L. and Mastorakis, N. (2009) Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579–588.

Mason, SJ, Goddard, L. 2001. Probabilistic precipitation anomalies associated with ENSO. *Bulletin of the American Meteorological Society*, 82(4): 619–638. https://doi.org/10.1175/1520-0477(2001)082<0619:PPAAWE>2.3.CO;2.

Michaelsen, J. (1987) Cross-validation in statistical climate forecast models. *Journal of Climate and Applied Meteorology*, 26, 1589–1600.

Min, Y.-M., Kryjov, V.N. and Oh, J.H. (2011) Probabilistic interpretation of regression-based downscaled seasonal ensemble predictions with the estimation of uncertainty. *Journal of Geophysical Research: Atmospheres*, 116(8), 1–10. https://doi.org/10.1029/2010JD015284.

Min, Y.-M., Kryjov, V.N. and Oh, S.M. (2014) Assessment of APCC multimodel ensemble prediction in seasonal climate forecasting: retrospective (1983-2003) and real-time forecasts (2008-2013). *Journal of Geophysical Research: Atmospheres*, 119 (12), 132–150. https://doi.org/10.1002/2014JD022230.

Min, Y.-M., Kryjov, V.N., Oh, S.M. and Lee, H.-J. (2017) Skill of real-time operational forecasts with the APCC multi-model ensemble prediction system during the period 2008–2015. *Climate Dynamics*, 49(11), 4141–4156. https://doi.org/10.1007/s00382-017-3576-2.

Min, Y.-M., Kryjov, V.N., Park, C.-K. (2009) A probabilistic multi-model ensemble approach to seasonal prediction. *Weather and Forecasting*, 24(3), 812–828. http://dx.doi.org/10.1175/2008waf2222140.1.

Mishra, N., Prodhomme, C. and Guemas, V. (2019) Multi-model skill assessment of seasonal temperature and precipitation forecasts over Europe. *Climate Dynamics*, 52(7–8), 4207–4225. https://doi.org/10.1007/s00382-018-4404-z.

Murphy, A.H. (1998) Skill scores based on the mean square error and their relationship to the correlation coefficient. *Monthly Weather Review*, 116(12), 2417–2424 https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Palmer, T.N., Doblas-Reyes, F.J., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., Feddersen, H., Cantelaube, P., Terres, J.-M., Davey, M., Graham, R., Délécluse, P., Lazar, A., Déqué, M., Guérémy, J.-F., Díez, E., Orfila, B., Hoshen, M., Morse, A.P., Keenlyside, N., Latif, M., Maisonnave, E., Rogel, P., Marletto, V. and Thomson, M.C. (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (Demeter). *Bulletin of the American*

Meteorological Society, 85(6), 853–872. https://doi.org/10.1175/BAMS-85-6-853.

Park, H.-J., Kryjov, V.N. and Ahn, J.-B. (2018) One-month-lead predictability of Asian summer monsoon indices based on the zonal winds by the APCC multimodel ensemble. *Journal of Climate*, 31(21), 8945–8960. https://doi.org/10.1175/JCLI-D-17-0816.1.

Park, S., Kang, S.-D. and Kwon, W.-T. (2005) Prediction of boreal winter precipitation by nonlinear multimodel ensemble technique (in Korean with English abstract). *Asia-Pacific Journal of Atmospheric Sciences*, 41(6), 1015–1028.

Peng, P., Kumar, A., Van Den Dool, H. and Barnston, A.G. (2002) An analysis of multimodel ensemble predictions for seasonal climate anomalies. *Journal of Geophysical Research: Atmospheres*, 107(23), 1–12. https://doi.org/10.1029/2002JD002712.

Peng, P., Kumar, A. and Wang, W. (2011) An analysis of seasonal predictability in coupled model forecasts. *Climate Dynamics*, 36 (3), 637–648. https://doi.org/10.1007/s00382-009-0711-8.

Ratnam, J.V., Dijkstra, H.A., Doi, T., Morioka, Y., Nonaka, M. and Behera, S.K. (2019) Improving seasonal forecasts of air temperature using a genetic algorithm. *Scientific Reports*, 9(1), 1–11. https://doi.org/10.1038/s41598-019-49281-z.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C. and Wang, W. (2002) An improved in situ and satellite SST analysis for climate. *Journal of Climate*, 15, 1609–1625 https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

Rodrigues, L.R.L., Doblas-Reyes, F.J. and Coelho CA dos, S. (2014) Multi-model calibration and combination of tropical seasonal sea surface temperature forecasts. *Climate Dynamics*, 42(3–4), 597–616. https://doi.org/10.1007/s00382-013-1779-8.

Ropelewski, C.F. and Halpert, M.S. (1989) Precipitation patterns associated with the high index phase of the southern oscillation. *Journal of Climate*, 2, 268–284 https://doi.org/10.1175/1520-0442(1989)002<0268:PPAWTH.

Saji, N.H., Goswami, P.N., Vinayachandran, P.N. and Yamagata, T. (1999) A dipole mode in the tropical Indian Ocean. *Nature*, 401 (September), 360–363.

Setiawan, A.M., Lee, W.S. and Rhee, J. (2017) Spatio-temporal characteristics of Indonesian drought related to El Niño events and its predictability using the multi-model ensemble. *International Journal of Climatology*, 37(13), 4700–4719. https://doi.org/10.1002/joc.5117.

Shin, S.-H. and Moon, J.-Y. (2018) Prediction skill for the East Asian winter monsoon based on APCC multi-models. *Atmosphere*, 9(8), 1–18. https://doi.org/10.3390/atmos9080300.

Shukla, J., Anderson, J., Baumhefner, D., Brankovic, C., Chang, Y., Kalnay, E., Marx, L., Palmer, T., Paolino, D., Ploshay, J., Schubert, S., Straus, D., Suarez, M. and Tribbia, J. (2000) Dynamical seasonal prediction. *Bulletin of the American Meteorological Society*, 81(11), 2593–2606 https://doi.org/10.1175/1520-0477(2000)081<2593:DSP>2.3.CO;2.

Sohn, S.-J. and Tam, C.-Y. (2016) Long-lead station-scale prediction of hydrological droughts in South Korea based on bivariate pattern-based downscaling. *Climate Dynamics*, 46(9–10), 3305–3321. https://doi.org/10.1007/s00382-015-2770-3.

Sohn, S.-J., Tam, C.-Y. and Jeong, H.-I. (2016) How do the strength and type of ENSO affect SST predictability in coupled models. *Scientific Reports*, 6(September), 1–8. https://doi.org/10.1038/srep33790.

Sohn, S.-J., Tam, C.-Y. and Park, C.-K. (2011) Leading modes of east Asian winter climate variability and their predictability: An assessment of the APCC multi-model ensemble. *Journal of the Meteorological Society of Japan*, 89(5), 455–474. https://doi.org/10.2151/jmsj.2011-504.

Stockdale, T. 2012. The EUROSIP System. *ECMWF Seminar on Seasonal Prediction*.

Stouffer, R.J., Manabe, S. and Vinnikov, K.Y. (1994) Model assessment of the role of natural variability in recent global warming. *Nature*, 367(6464), 634–636. https://doi.org/10.1038/367634a0.

Taylor, K.E. (2005) Taylor Diagram Primer. http://www.atmos.albany.edu/daes/atmclasses/atm401/spring_2016/ppts_pdfs/Taylor_diagram_primer.pdf.

Trenberth, K.E., Branstator, G.W., Karoly, D., Kumar, A., Lau, N.-C. and Ropelewski, C. (1998) Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research: Oceans*, 103(C7), 14291–14324. https://doi.org/10.1029/97jc01444.

Trenberth, K.E. and Paolino, D.A. (1981) Characteristic patterns of variability of sea level pressure in the Northern Hemisphere. *Monthly Weather Review*, 109(6), 1169–1189 https://doi.org/10.1175/1520-0493(1981)109<1169:CPOVOS>2.0.CO;2.

Wang, B., Lee, J.-Y., Kang, I.-S., Shukla, J., Park, C.-K., Kumar, A., Schemm, J., Cocke, S., Kug, J.-S., Luo, J.-J., Zhou, T., Wang, B., Fu, X., Yun, W.-T., Alves, O., Jin, E.K., Kinter, J., Kirtman, B., Krishnamurti, T., Lau, N.C., Lau, W., Liu, P., Pegion, P., Rosati, T., Schubert, S., Stern, W., Suarez, M. and Yamagata, T. (2009) Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980-2004). *Climate Dynamics*, 33(1), 93–117. https://doi.org/10.1007/s00382-008-0460-0.

Wilks, D.S. (1995) *Statistical Methods in the Atmospheric Sciences: An Introduction*. San Diego: Elsevier Science.

WMO. (2007) *Report of WMO/KMA GPCs Workshop on Lead Centre for Long-Range Forecast Multi-Model Ensembles (LC-LRFMME)*. Busan: Republic of Korea Available at: http://www.wmo.int/pages/prog/www/DPFS/Reports/Wshop-LCLRFMME_Busan.doc.

WMO. (2008) *Meeting of Expert Team on Extended and Long-Range Forecasting*. China: Beijing Available at: http://www.wmo.int/pages/prog/www/DPFS/Reports/ET-ELRF_Beijing2008.doc.

WMO. 2010. Standardized verification system (SVS) for long-range forecasts (LRF), attachment II.8.1-II.8.17 to the manual on the GDPS. WMO-No. 485.

WMO. (2016) *WMO Regional Climate Outlook Forums*. Geneva: Switzerland Available at: http://library.wmo.int/opac/doc_num.php?explnum_id=3191.

WMO. 2018. Manual on the global data-processing and forecasting system (annex IV to the WMO technical regulations) 2017 edition updated in 2018. WMO-No. 485.

You, S.H., Lee, Y.H. and Lee, W.J. (2012) Parameter estimations of a storm surge model using a genetic algorithm. *Natural Hazards*, 60(3), 1157–1165. https://doi.org/10.1007/s11069-011-9900-y.

Yun, W.-T. and Krishnamurti, T.N. (2002) Linear and non-linear multi-model superensemble prediction model. *Atmosphere*, 12 (3), 26–31.

Yun, W-T, Stefanova, L, Krishnamurti, TN. 2003. Improvement of the multimodel superensemble technique for seasonal forecasts. *Journal of Climate*, 16(22): 3834–3840. https://doi.org/10.1175/1520-0442(2003)016<3834:IOTMST>2.0.CO;2.

Zadra, A., Williams, K., Frassoni, A., Rixen, M., Adames, Á.F., Berner, J., Bouyssel, F., Casati, B.B., Christensen, H., Ek, M.B., Flato, G., Huang, Y., Judt, F.K., Lin, H., Maloney, E., Merryfield, W., Van Niekerk, A., Rackow, T., Saito, K.Z., Wedi, N. and Yadav, P.K. (2018) Systematic errors in weather and climate models: nature, origins, and ways forward. *Bulletin of the American Meteorological Society*, 99(4), ES67–ES70. https://doi.org/10.1175/BAMS-D-17-0287.1.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.