

Statistical downscaling methods based on APCC multi-model ensemble for seasonal prediction over South Korea

Suchul Kang,^a Jina Hur^b and Joong-Bae Ahn^{b*}

^a *Climate Research Department, APEC Climate Center, Busan, Republic of Korea*

^b *Division of Earth Environmental System, Pusan National University, Busan, Republic of Korea*

ABSTRACT: An investigation was conducted to optimize the application of the multi-model ensemble (MME) technique for statistical downscaling using 1- to 6-month lead hindcasts obtained from six operational coupled general circulation models (GCMs) participating in the APEC Climate Center (APCC) MME prediction system. Three different statistical downscaling MME methods (SDMMEs) were compared and estimated over South Korea. The study results revealed that under the same number of ensemble members, simple changes in the statistical downscaling method, such as an applicative order or a type of MME, can help to improve the predictability. The first method, the conventional technique, performed MME using data downscaled from the single-model ensemble means of each GCM (SDMME-Sm), whereas the second and third methods, newly designed in this study, calculated the simple ensemble mean (SDMME-Ae) and the weighted ensemble mean (SDMME-We) after statistical downscaling for each member of all model ensembles. These three methods were applied to predict temperature and precipitation for the 6-month summer-fall season over 23 years (1983–2005) at 60 weather stations over South Korea. The predictors were variables from hindcasts integrated by the six coupled GCMs. According to the analysis, both SDMME-Ae and SDMME-We showed increased predictability compared with SDMME-Sm. In particular, SDMME-We showed more significant improvement in long-term prediction. In addition, in order to assess the dependence of predictability on the number of downscaled ensemble members and the type of MME, an additional experiment was performed, the results of which revealed that the model performance was closely related to the number of downscaled ensemble members. However, after approximately 30 ensemble members, the predictive skills became rapidly saturated when using the SDMME-Ae method. SDMME-We overcame the limited skills that can be achieved by merely increasing the number of downscaled ensemble members, thereby improving the performance.

KEY WORDS statistical downscaling; seasonal prediction; multi-model ensemble; regional climate; APCC MME prediction

Received 6 July 2013; Revised 19 December 2013; Accepted 17 January 2014

1. Introduction

In the past decade, multi-model ensemble (MME) techniques have been developed to improve the accuracy in seasonal predictions by reducing uncertainties associated with individual models (Krishnamurti *et al.*, 1999; Peng *et al.*, 2002; Yun *et al.*, 2003; Palmer *et al.*, 2004). Much of the effort in developing MME methods has been devoted to producing skilful seasonal forecasts, in order to provide users with relatively reliable long-term information using global climate models (e.g. Krishnamurti *et al.*, 1999; Palmer and Shukla, 2000). Although the MME prediction based on global climate models is useful in many respects, it remains incapable of being used for agricultural or hydrological purposes, which require station-based or high-resolution data, due to the low-resolution grid system of the models with an approximate horizontal dimension of 100–300 km.

Recently, dynamical and statistical downscaling techniques have been rapidly developing as a method of

overcoming the constraints of the global climate prediction model, thereby offering high-resolution data (Kang *et al.*, 2009; Vasiliades *et al.*, 2009; Ahn *et al.*, 2012). Dynamical downscaling using a regional climate model produces physically and dynamically balanced data. However, it has shortcomings in composing the MME system because of the enormous computing time and storage capacity for each ensemble member (Chen *et al.*, 2012) and the systematic model bias (Ahn *et al.*, 2012). Therefore, statistical downscaling techniques are often used to effectively predict the station-based climate by taking the MME members into consideration and using model output statistics. The results derived from the statistical downscaling may retain considerable uncertainties because of systematic errors in the global climate model output (Glahn and Lowry, 1972; Wilks, 1995). By applying the MME scheme to predictors or predictands, however, these uncertainties can be partially reduced (Kang *et al.*, 2009; Juneng *et al.*, 2010).

The Asia-Pacific Economic Cooperation Climate Center (APCC) has been operating an MME seasonal forecast system by collecting 6-month lead predictions from the world's leading research institutions since 2005 (Min *et al.*, 2009). They have derived not only global-scale MME seasonal predictions from general circulation

* Correspondence to: J.-B. Ahn, Department of Atmospheric Sciences, Pusan National University, Busan 609735, Republic of Korea. E-mail: jbahn@pusan.ac.kr

Table 1. Description of six GCMs used as predictors.

Model acronym	Institution(country)	Model resolution	Ensemble size	References
APCC	APEC Climate Center (Korea)	T85L26	5	Jeong <i>et al.</i> (2008)
NCEP	NCEP Climate Prediction Center (USA)	T62L64	15	Saha <i>et al.</i> (2006)
PNU	Pusan National University (Korea)	T42L18	5	Sun and Ahn (2011)
POAMA	Bureau of Meteorology Research Centre (Australia)	T47L17	10	Zhao and Hendon (2009)
SNU	Seoul National University (Korea)	T42L21	6	Ham and Kang (2011)
UH	University of Hawaii (USA)	T31L19	10	Fu and Wang (2004)

models (GCMs), but also regional forecasts using statistical downscaling models (Kang *et al.*, 2007; Chu *et al.*, 2008; Kang *et al.*, 2009; Min *et al.*, 2011). Kang *et al.* (2009) have shown that 3-month lead forecasts based on the GCM results from APCC can be skilfully used to predict local variables over South Korea by using an appropriate statistical downscaling method. Subsequently, Min *et al.* (2011) suggested a modified procedure of the predictor selection using basically the same model in order to account for the physical relationships between predictors and predictands. Given that a large ensemble size can lead to skill improvement (Palmer *et al.*, 2000; Palmer *et al.*, 2004), however, their experiment downscaled by a single-model ensemble (SME) retains limited skill due to the restrictive ensemble size.

APCC recently developed a 6-month lead MME prediction system based on the prediction information produced by ocean–atmosphere coupled models through cooperation with six research institutes (Table 1). An optimal statistical downscaling method based on the MME approach in terms of station-based 6-month lead summer–fall (June to November) precipitation and temperature predictions is investigated using the data. Considering the characteristics of individual members, two statistical downscaling methods are newly designed and explored under the statistical downscaling framework of APCC by evaluating and comparing the result with the previous one (e.g. Kang *et al.*, 2009; Juneng *et al.*, 2010; Min *et al.*, 2011).

This paper is organized as follows. In Section 2, we present the use of datasets and statistical downscaling methods, including those newly designed for this study. The results and conclusions are illustrated in Sections 3 and 4, respectively.

2. Data and methodology

2.1. Data

The predictands of statistical downscaling are surface temperature and precipitation of the boreal summer–fall season (June to November) located at 60 weather stations over South Korea. To construct the downscaling model and verify the predictions, we used the observed precipitation and temperature data for the 23 years from 1983 to 2005, as provided by the Korea Meteorological Administration. Figure 1 shows the locations of the 60 weather stations as well as the topography of South Korea.

The predictors of the statistical model are multiple variables derived from 6-month lead hindcasts for summer–fall

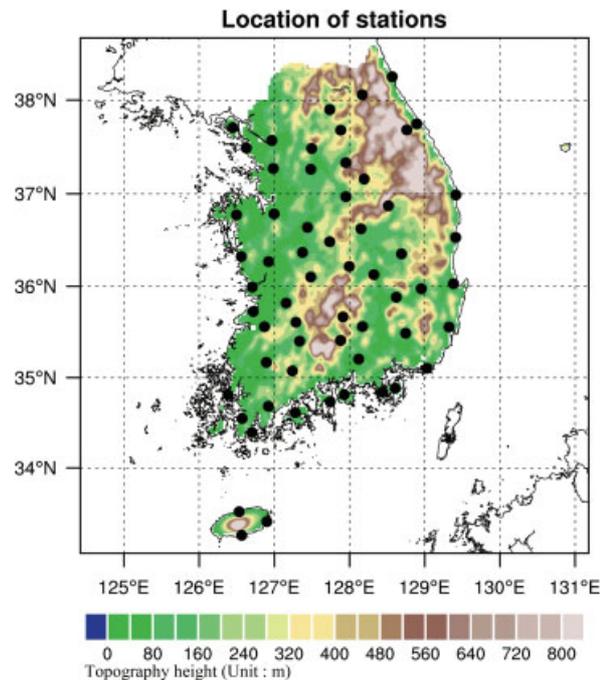


Figure 1. Locations of 60 weather stations (black dot) and topography (shaded, m) of South Korea.

produced by six operational coupled GCMs, participating in the APCC 6-month lead MME prediction system. The 23-year hindcast period (1983–2005), for which the APCC archive is available, is used in this study. The individual model is described in Table 1.

As in the study by Min *et al.* (2011), anomaly fields of 500-hPa geopotential height (Z500), sea level pressure (SLP) and 850-hPa temperature (T850) are used as potential predictors for temperature, while predictors for the precipitation are comprised of the SLP and Z500 anomalies. These potential predictors are selected to minimize the artificial skill in choosing the predictor, so-called screening (Delsole and Shukla, 2009), by considering the feasible physical relationships between the predictors and predictands (Kang and Shukla, 2006; Min *et al.*, 2011).

2.2. Methodology

Our statistical downscaling approach is comprised of two major steps. The first step is a screening procedure to select the optimal predictors. In this stage, a moving window technique, one of the pattern projection methods, is used (e.g. Kug *et al.*, 2008a; Kang *et al.*, 2009; Min *et al.*, 2011).

In this method, the area over the whole globe is scanned to find the region highly related to the predictand using a movable window with a size of 15×10 latitude–longitude grid points. A movable window with the largest sum of correlation coefficients between the predictand and the predictor is identified, and then the projection of predictors within this optimal window is obtained. In the second step, a statistical model is constructed and the target variables are predicted using the selected predictors. The linear relationship between the target variable and the predictor projection is estimated and modeled using simple linear regression (Juneng *et al.*, 2010). In estimating the optimal window and the regression coefficient, the relationship between each point of predictand and each ensemble member from the individual model is considered independently. This means that the optimal window and the regression coefficient are calculated 3060 times (a total of 60 predictand points \times 51 of the ensemble size) for forecasting a certain season. A double cross-validation method is applied to the statistical downscaled results constructed by the previous steps, in order to avoid overfitting of random noise, which is a common problem for all empirical prediction models (Yu *et al.*, 1997; Feddersen *et al.*, 1999; Juneng *et al.*, 2010; Min *et al.*, 2011). A detailed description of the model is presented in Min *et al.* (2011) and Kang *et al.* (2009).

The following three different statistical downscaling methods are designed on the basis of the same statistical model, and the results are compared to explore the appropriate statistical downscaling MME method. In the experimental design, the conventional MME methods are adopted because this study is focused on how to improve the predictive skill by combing the MME method and statistical downscaling instead of developing the MME method or the downscaling model.

- SDMME-Sm: The conventional technique, used in previous studies, estimates the MME mean using data downscaled from the SME means of the coupled GCMs (Kang *et al.*, 2009; Min *et al.*, 2011) (Figure 2(a)).
- SDMME-Ae: Newly designed in this study, which calculates the MME mean by simply averaging all model ensembles applied to statistical downscaling (Figure 2(b)).
- SDMME-We: The newly designed SDMME-We is similar to SDMME-Ae, but uses a weighted ensemble scheme instead of the simple average method for MME mean (Figure 2(c)). In this method, the members with a correlation coefficient greater than a threshold number at each point are chosen after performing cross-validation. To estimate the suitable threshold number, we performed an empirical test of whether at least one ensemble member from each model has a temporal correlation coefficient (TCC) greater than the threshold number for 1–6 month lead temperature and precipitation hindcasts. In selecting the proper criterion, we set 0.05 intervals for the threshold number from 0.20 to 0.50, and calculate the number of the

ensemble members having a TCC greater than each threshold (Table 2). According to the results, more than one ensemble member in each model can be involved in calculating SDMME-We when the threshold number is 0.3. Therefore, we choose TCC of 0.3 as the criterion for the SDMME-We method. Then the MME mean with a weighting is calculated as follows:

$$\text{SDMME} - \text{We} = \sum_{j=1}^n \left(\frac{\text{Corr}_j}{\sum_{i=1}^n \text{Corr}_i} \times F_j \right) / n$$

where, the number of members with correlation coefficients greater than 0.3 is defined as n , and the correlation coefficient and forecast of the i th member are indicated as Corr_i and F_j , respectively. The members having poor skill are excluded in the method rather than excluding the models of poor skill. This approach is taken in order to consider as many climate models as possible, because the individual models produce their own climate signal. Thus, at least one ensemble member per the individual model is considered in MME. However, it does not indicate that the each model should have the same weighting.

The previous studies (e.g. Kang *et al.*, 2009; Min *et al.*, 2011) have shown that the statistically downscaled hindcasts (SDMME) show a relatively better performance than the results from GCMs simply interpolated (IMME) to *in situ* observation sites. Consequently, this study focused on evaluating the predictabilities of the SDMME hindcasts with the three methods rather than comparing between the results from IMME and SDMME. The performance of the MME mean using data simply interpolated from the SME means (IMME-Sm) is also evaluated only briefly.

3. Results and discussions

First, the interannual variabilities of the downscaled temperature and precipitation hindcasts derived from the three methods are analysed. Furthermore, we compare the skill of SDMME hindcasts with that of IMME hindcasts in order to briefly show the performance of the raw model output. Figure 3 shows the distributions of the temporal correlation coefficients (TCCs) for downscaled temperature and precipitation by the three different SDMMEs and the IMME during summer at the 60 stations. In this figure, TCCs of 0.37, 0.43 and 0.55 indicate 90, 95 and 99% confidence levels, respectively. IMME-Sm has TCCs lower than 0.37 at all stations with an average of -0.1 for 3-month lead precipitation, which is not statistically significant. However, SDMME-Sm shows positive TCCs with an average of 0.4 at 55 of the stations. The average TCC for SDMME-Ae over all the stations is 0.56, which is significant at the 99% confidence level. Considerable improvement is achieved when SDMME-We is compared with SDMME-Sm and SDMME-Ae in terms of TCC, although

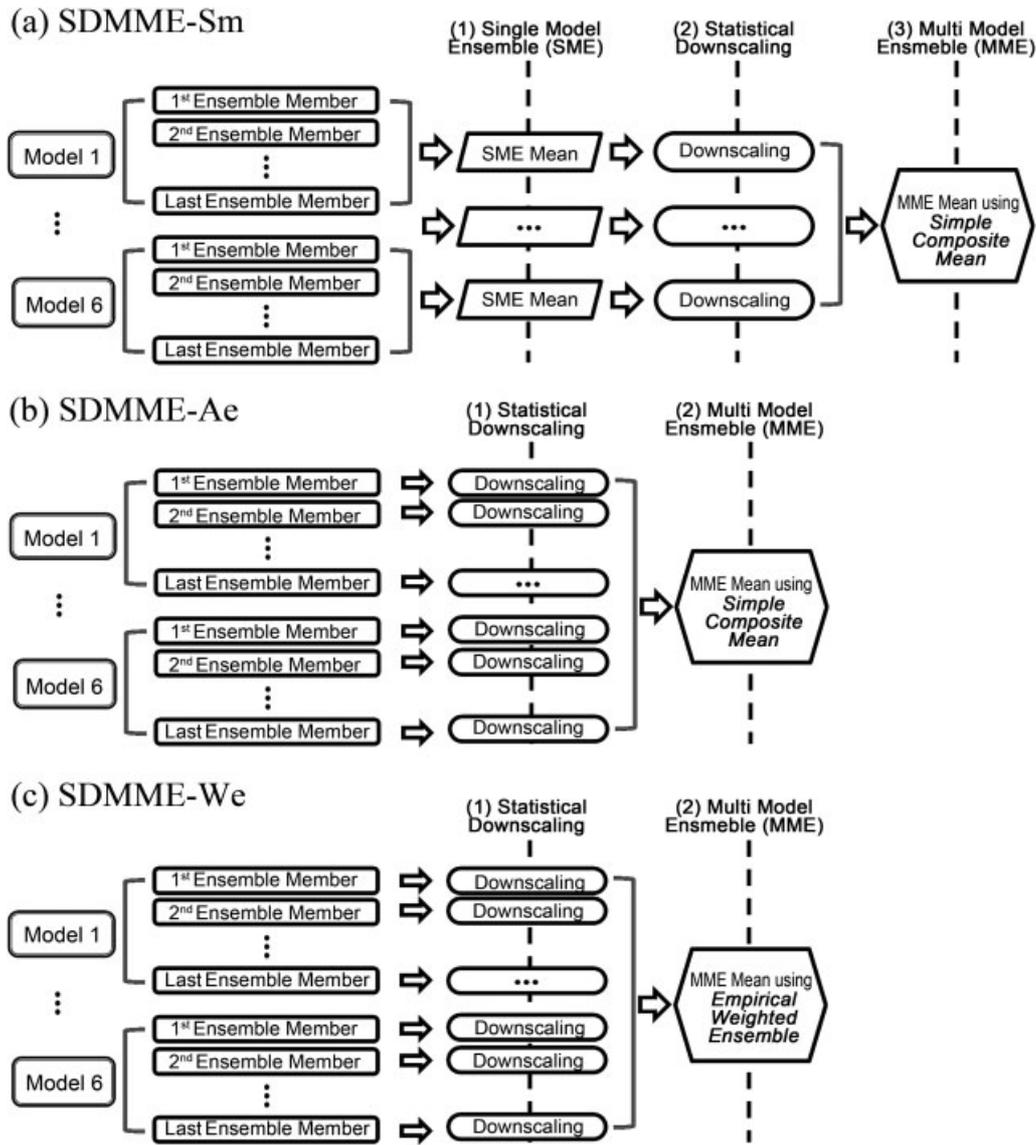


Figure 2. Schematic diagrams of the three experiments performed.

SDMME-Ae is already sufficiently higher than the 95% confidence level at most of the stations. This indicates that the precipitations predicted with SDMME-We are statistically significant at almost all of the stations, while SDMME-Sm has less skill compared with SDMME-We and SDMME-Ae.

As in the case of precipitation, SDMME-We also shows the highest TCC for temperature, followed in order by SDMME-Ae, SDMME-Sm and IMME-Sm. TCC derived from IMME-Sm for temperature also reveals the lowest value with an average of -0.09 compared with that from SDMMEs. The low skill of IMME can be attributed to the difficulty experienced by the current dynamical models in predicting a local variability correctly at each grid point, although they are capable of capturing a large-scale pattern related to a local variability over the continental region and extratropical oceans (Kug *et al.*, 2008b). The differences in averaged TCCs between SDMME-We and SDMME-Sm are 0.37 and 0.36 for precipitation and

Table 2. The number of models having at least one ensemble member with temporal correlation coefficient (TCC) greater than each threshold number.

Season	JJA (1–3 month lead)		SON (4–6 month lead)	
	PREC	TEMP	PREC	TEMP
Threshold number				
0.20	6	6	6	6
0.25	6	6	6	6
0.30	6	6	6	6
0.35	6	5	6	5
0.40	6	5	6	5
0.45	6	5	6	5
0.50	4	4	6	5

temperature, respectively, which are statistically significant at the 99% confidence level based on Fisher's transformation. This implies that the predictability decreases considerably for both temperature and precipitation when

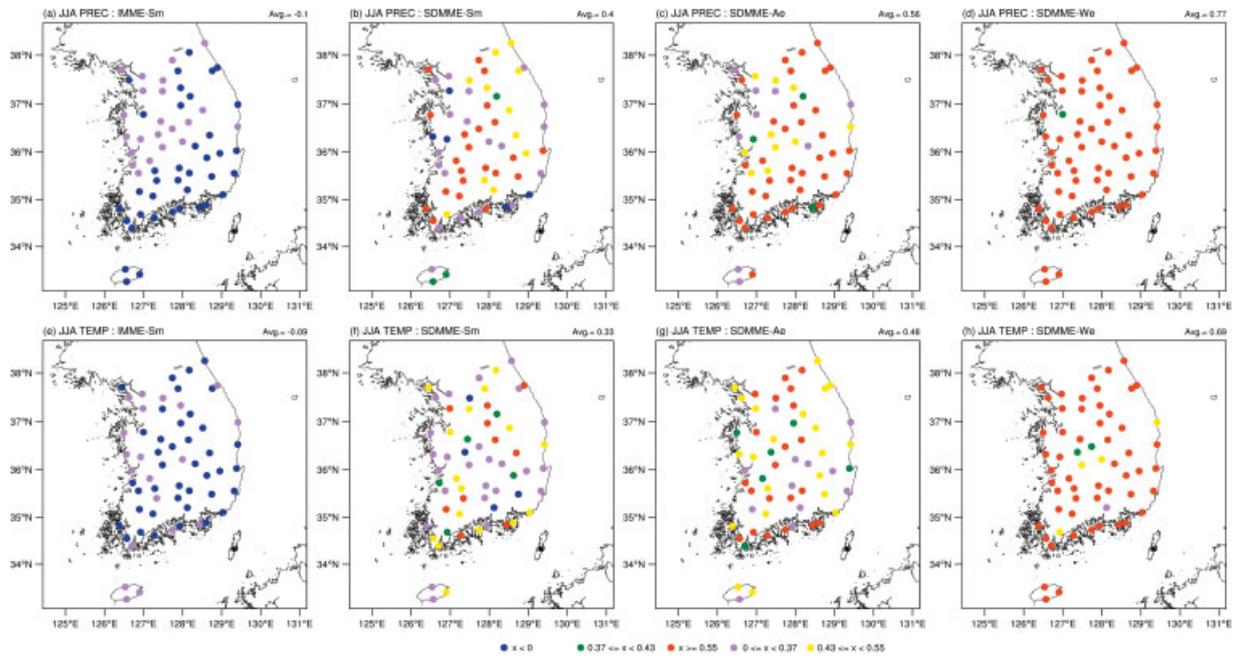


Figure 3. Temporal correlation coefficients (TCCs) for 3-month precipitation (PREC; upper panel) and temperature (TEMP; lower panel) hindcasts in JJA at each station by applying IMME-Sm (the first column), SDMME-Sm (the second column), SDMME-Ae (the third column) and SDMME-We (the last column).

the information from individual ensemble members is disregarded during MME. In other words, the downscaled temperature and precipitation hindcasts are sensitive to the characteristics of each ensemble member.

We also tried to determine whether these techniques are significant up to 4- to 6-month lead hindcasts, as well as 1- to 3-month. Figure 4 illustrates the monthly mean TCCs for 1- to 6-month lead temperature and precipitation hindcasts. The TCCs are calculated at each station and then averaged for all stations. The TCCs of SDMME-Sm are lower than those of SDMME-Ae and SDMME-We and more variable for both temperature and precipitation throughout the entire lead period. This implies that SDMME-Sm is more sensitive to individual information contained in the downscaled SME data because the number of MME members is relatively fewer than that of both SDMME-Ae and SDMME-We. Thus the TCCs of SDMME-Ae and SDMME-We show a more stable pattern with a steady decrease as the lead month increases. Moreover, SDMME-Ae and SDMME-We provide a sufficient skill improvement compared with SDMME-Sm for each lead period, as shown in Figure 4. In particular, the performance of SDMME-We is remarkably improved in prediction for both temperature and precipitation with a 99% confidence level throughout all the leads.

An ensemble spread can be used as an indicator to measure the predictability or uncertainty in issuing a deterministic forecast (WMO, 2012). In order to examine the reason why SDMME with all model ensembles has good performance, we also calculated the mean spread of the ensembles used in SDMME-Sm, and SDMME-Ae for precipitation and temperature hindcasts in summer and fall (Table 3). It shows that the spread of SDMME-Ae is

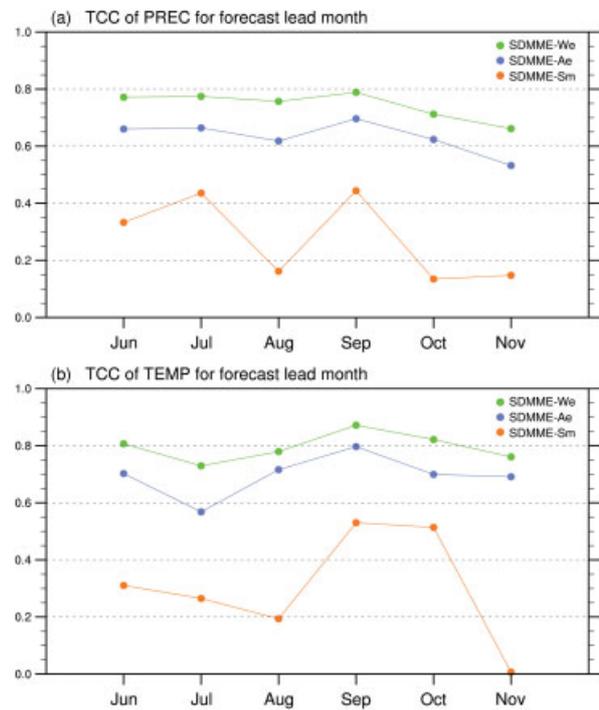


Figure 4. Monthly temporal correlation coefficients (TCCs) of (a) precipitation (PREC) and (b) temperature (TEMP) anomalies for 6-month lead (June to November) hindcasts.

lower than that of SDMME-Sm, especially in precipitation for JJA (June–August). Considering that a large spread generally indicates low predictability, SDMME-Ae provides superior capability in seasonal forecasts compared with SDMME-Sm. The low spread of SDMME-Ae arises

ensemble member is important because it offers additional information apart from the ensemble mean, although they are from the same model. Consequently, the use of all members enables SDMME-Ae and SDMME-We to offer better predictability by considering more of the large-scale information, compared with SDMME-Sm.

We conducted an additional experiment using the first five ensembles per model with SDMME-Ae and SDMME-We in order to inspect the degree of contribution of the different ensemble sizes from each GCM. In this experiment, we do not consider SDMME-Sm because the different ensemble size is less important in the method. Figure 5 compares the performances obtained with the different ensemble size depending on the model and the same number of ensemble members. The TCC derived from the five members is slightly lower than that from all members but still quite high. This difference may have been due to the reduced total ensemble size, or the removal of the implicit weight of the models. The total number of ensemble members is 30 when only five members are chosen, which is smaller than the ensemble size of 51 from all members. However, 30 members are enough to improve the skill (more described in Figure 7), which in turn does not lead to a conspicuous increase of predictability. In the model weighting terms, the impact of three models, NCEP, POAMA and UH, having more than ten ensemble members is reduced when the same ensemble size from the individual models is used. Despite their sufficiently good ability in long-term forecasting, these models did not exhibit any superiority over the others in the SME predictability (Figure 6). The different ensemble size from each GCM might increase the predictive capability compared with that obtained having the same number of ensemble members per model, but it does not cause a great discrepancy.

To illustrate the performance of the downscaled GCM outputs for each method, the TCCs between the downscaled GCM outputs and the observation at each station are calculated and then averaged for the 60 stations (Figure 6). Figure 6 shows that the predictive ability generally decreases in the order of SDMME-We, SDMME-Ae and SDMME-Sm. Especially, SDMME-We generates the most pronounced predictability from the limited information of the individual GCM outputs. This is similar with the results of previous studies (e.g., Robertson *et al.*, 2004; Yun *et al.*, 2005; Kug *et al.*, 2008b) that used the empirical-weighted MME. Kug *et al.* (2008b) attributed the result to the skill dependence on the past performance of the individual member. If the predictive skill of individual member exceeds a certain level, the weighted MME method would not significantly improve predictability. However, seasonal predictions have relatively poor skills over continental regions and extratropics. Therefore, the weighted MME method using only reliable predictions has better performance than that of SCM. Furthermore, the skill of MME is the highest, regardless of the method. This agrees with the claims of Leith (1974), Doblas-Reyes *et al.* (2000) and Palmer *et al.* (2000) that the predictive skill of the ensemble mean is, on average, relatively higher than

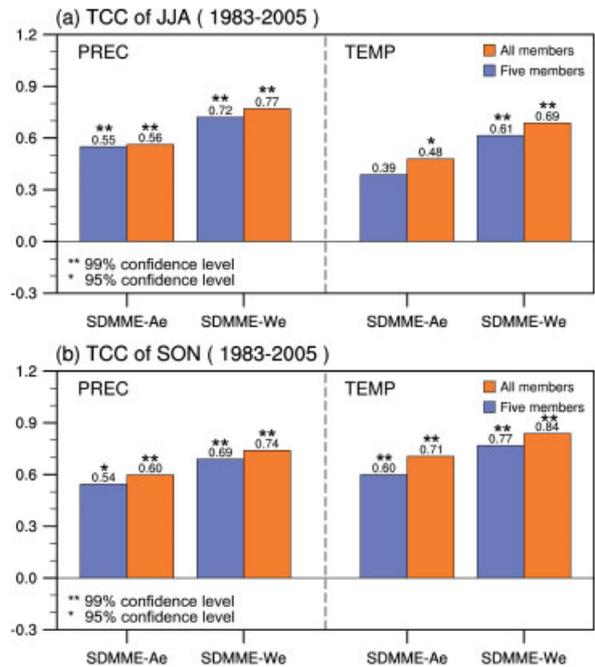


Figure 5. Comparison between the temporal correlation coefficients (TCCs) derived from all ensemble members and the first five members per each model for 6-month precipitation (PREC) and temperature (TEMP) hindcasts.

that of any individual member because these MME methods might reduce some of the systematic bias.

To explain the superior predictability of SDMME-We over SDMME-Ae and SDMME-Sm, we investigate the sensitivities of predictabilities to the number of downscaled ensemble members. First, we conduct MME experiments by changing the number of statistically downscaled ensemble members for both seasons. MME means are calculated using simple composite averages by increasing the number of ensemble members from 5 to 50. The members of each ensemble are selected from 51 total ensemble members in Table 1. Each ensemble having the same number of members is randomly produced 1000 times. Then, the TCC means are calculated for each and every combination. Figure 7 shows that TCC increases in proportion to the increase in the number of downscaled ensemble members. According to the results, temperature is particularly sensitive to the chosen members and the TCCs are rapidly saturated after the number of ensemble members exceeds approximately 30, in agreement with Palmer *et al.* (2004). The mean TCC for SDMME-We is higher than any other results, as illustrated in the previous figures, and almost corresponds to the SDMME-Ae TCC when 50 ensemble members are used. This implies that predictability depends on the number of downscaled ensemble members in the SCM, until they number approximately 30, and that the weighted MME method can complement the limitation of obtainable predictability by increasing the number of downscaled ensemble members. Although the mere increase in the number of ensemble members increases the predictability of the model (Palmer *et al.*, 2000; Palmer *et al.*, 2004; Robertson *et al.*, 2004), more studies are

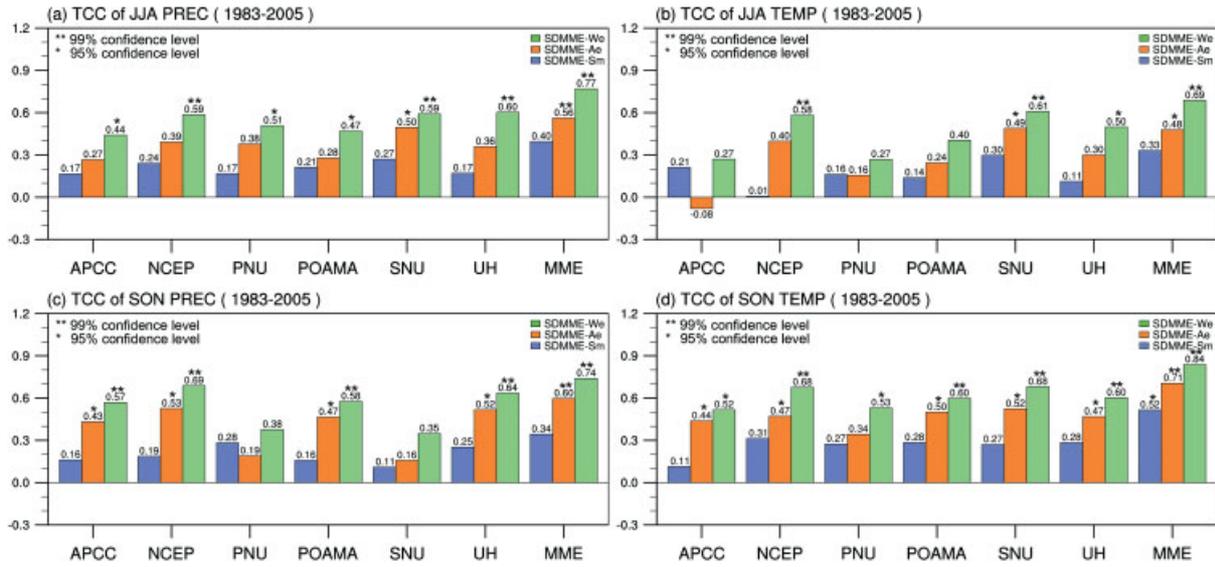


Figure 6. Temporal correlation coefficients (TCCs) between the downscaled GCM output used in the three methods and the *in situ* observations for precipitation (PREC; left column) and temperature (TEMP; right column) anomalies in JJA (1–3 month lead; upper panel) and SON (4–6 month lead; lower panel), averaged over all stations.

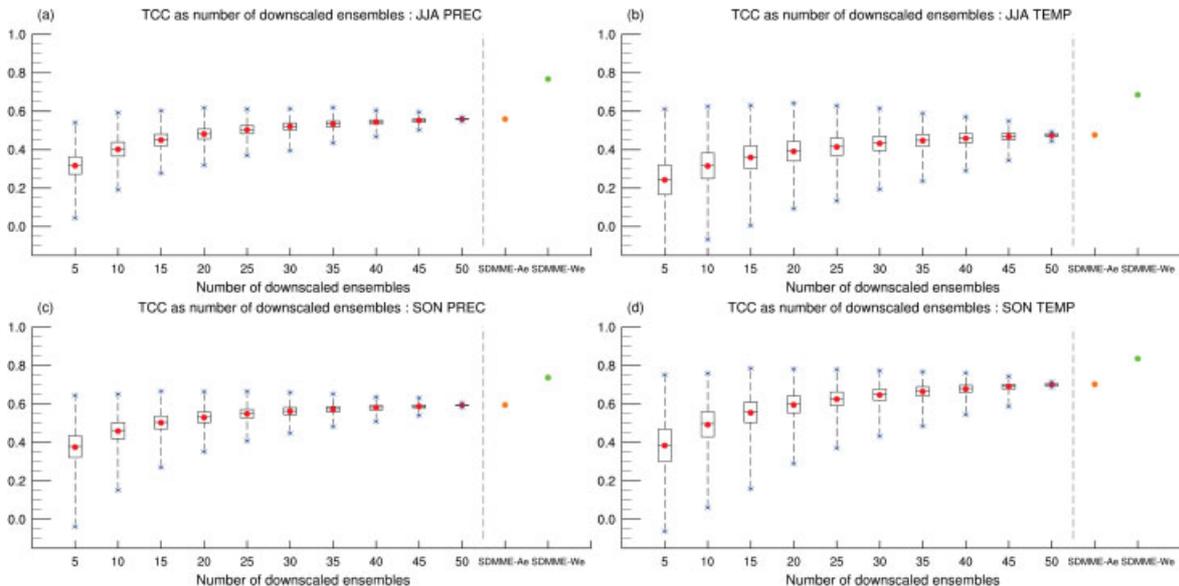


Figure 7. Changes in temporal correlation coefficients (TCCs) of precipitation (PREC; left column) and temperature (TEMP; right column) hindcasts for JJA (1–3 month lead; upper panel) and SON (4–6 month lead; lower panel) as the number of downscaled ensemble members is increased from 5 to 50. The TCCs are calculated for each and every MME result using the SDMME-Ae method.

needed to explain the positive relationship between the downscaled ensemble size and predictability.

4. Concluding remarks

In this study, we estimated and compared the accuracy of three different statistical downscaling methods, SDMME-Sm, SDMME-Ae and SDMME-We, by applying them to 6-month lead hindcasts from 6 coupled GCMs for the optimal operational use of the APCC regional long-range MME prediction system. For the experiment, we produced and analysed the statistically downscaled

summer-fall temperature and precipitation hindcasted by the coupled GCMs for the 23 years from 1983 to 2005 for 60 *in situ* weather stations over South Korea.

According to the analysis, SDMME-We showed the best performance in terms of temporal and spatial patterns of the observation. According to the quantitative and categorical estimations of TCC, PCC, RMSE, HSS, HR and FAR, the SDMME-We predictability was statistically significant up to a lead-time of 4–6 months. This SDMME-We predictability appeared at the downscaled individual GCM output as well as the MME mean, which reveals the good predictability offered by SDMME-We from the limited

information of the individual GCM output. Furthermore, the skill of MME was superior to that of any single individual prediction.

Through additional MME experiments, we observed that the increased number of downscaled ensemble members improved the predictive skill. Although previous studies (e.g. Palmer *et al.*, 2000; Palmer *et al.*, 2004; Robertson *et al.*, 2004) already referred to the positive relationship between the number of ensemble members and the skills in MME with GCM output, we found that the increased predictability became saturated after approximately 30 downscaled ensemble members. Despite this limitation in the predictability improvement gained through increasing the number of downscaled ensemble members, we suggest that the weighted MME technique can be used to complement this limitation. This result supports the contention of previous studies (e.g. Robertson *et al.*, 2004; Yun *et al.*, 2005; Kug *et al.*, 2008b) that the empirical-weighted MME can induce skilful prediction. Further studies will be necessary to determine how the increased number of downscaled ensemble members triggers the skill improvement.

Our study reveals that simple changes, such the applicative order or MME method, can improve the predictability considerably, under the same number of ensemble members from GCMs. This result is applicable to dynamical downscaling methods based on MME.

Although double cross-validation is employed in this study, as highlighted by Delsole and Shukla (2009), even good performance in a model's cross-validation mode does not necessarily guarantee that it will perform well in real-time forecasts. This reveals the need for future study to demonstrate the actual usefulness of this approach using independent datasets with substantial long-period.

Acknowledgements

This work was carried out with the support of the Korea Meteorological Administration Research and Development Program under Grant CATER 2012–3100 and Rural Development Administration Cooperative Research Program for Agriculture Science and Technology Development under Grant Project No. PJ009353, Republic of Korea.

References

Ahn J-B, Lee J, Im E-S. 2012. The reproducibility of surface air temperature over South Korea using dynamical downscaling and statistical correction. *J. Meteorol. Soc. Jpn.* **90**(4): 493–507, DOI: 10.2151/jmsj.2012-404.

Chen H, Xu C-Y, Guo S. 2012. Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff. *J. Hydrol.* **434–435**: 36–45, DOI: 10.1016/j.jhydrol.2012.02.040.

Chu J-L, Kang H, Tam C-Y, Park C-K, Chen C-T. 2008. Seasonal forecast for local precipitation over northern Taiwan using statistical downscaling. *J. Geophys. Res.* **113**: D12118, DOI: 10.1029/2007JD009424.

Delsole T, Shukla J. 2009. Artificial skill due to predictor screening. *J. Clim.* **22**: 331–345, DOI: 10.1175/2008JCLI2414.1.

Doblas-Reyes FJ, Déqué M, Pielieuvre JP. 2000. Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Q. J. R. Meteorol. Soc.* **126**: 2069–2087, DOI: 10.1002/qj.49712656705.

Fedderson H, Navarra A, Ward MN. 1999. Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Clim.* **12**: 1974–1989, DOI: 10.1175/1520-0442(1999)012<1974:ROMSEB>2.0.CO;2.

Fu X, Wang B. 2004. The boreal-summer intraseasonal oscillations simulated in a hybrid coupled atmosphere–ocean model. *Mon. Weather Rev.* **132**: 2628–2649, DOI: 10.1175/MWR2811.1.

Glahn HR, Lowry DA. 1972. The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteorol.* **11**: 1203–1211, DOI: 10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2.

Ham YG, Kang I-S. 2011. Improvement of seasonal forecasts with inclusion of tropical instability waves on initial conditions. *Clim. Dyn.* **36**: 1277–1290, DOI: 10.1007/s00382-010-0743-0.

Jeong H-I, Ashok K, Song B-G, Min Y-M. 2008. Experimental 6-month hindcast and forecast simulation using CCSM3, APCC 2008 Technical Report, APEC Climate Center: Busan, South Korea.

Juneng L, Tangang FT, Kang H, Lee W-J, Seng YK. 2010. Statistical downscaling forecasts for winter monsoon precipitation in Malaysia using multimodel output variables. *J. Clim.* **23**: 17–27, DOI: 10.1175/2009JCLI2873.1.

Kang I-S, Shukla J. 2006. In *Dynamic Seasonal Prediction and Predictability of the Monsoon*, Wang B (ed). Praxis, DOI: 10.1007/3-540-37722-0_15; Chichester, UK; 585–612.

Kang H, An K-H, Park C-K, Solis ALS, Stithichivapak K. 2007. Multimodel output statistical downscaling prediction of precipitation in the Philippines and Thailand. *Geophys. Res. Lett.* **34**: L15710, DOI: 10.1029/2007GL030730.

Kang H, Park C-K, Hameed SN, Ashok K. 2009. Statistical downscaling of precipitation in Korea using multimodel output variables as predictors. *Mon. Weather Rev.* **137**: 1928–1938, DOI: 10.1175/2008MWR2706.1.

Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, Gadgil S, Surendran S. 1999. Improved weather and seasonal climate forecasts from multimodel superensemble. *Science* **285**: 1548–1550, DOI: 10.1126/science.285.5433.1548.

Kug J-S, Lee J-Y, Kang I-S. 2008a. Systematic error correction of dynamical seasonal prediction of sea surface temperature using a stepwise pattern project method. *Mon. Weather Rev.* **136**: 3501–3512, DOI: 10.1175/2008MWR2272.1.

Kug J-S, Lee J-Y, Wang B, Park C-K. 2008b. Optimal multi-model ensemble method in seasonal climate prediction. *Asia Pac. J. Atmos. Sci.* **44**: 259–267.

Leith CE. 1974. Theoretical skill of Monte Carlo forecasts. *Mon. Weather Rev.* **102**: 409–418.

Min Y-M, Kryjov VN, Park C-K. 2009. A probabilistic multimodel ensemble approach to seasonal prediction. *Weather Forecast* **24**: 812–828.

Min Y-M, Kryjov VN, Oh J-H. 2011. Probabilistic interpretation of regression-based downscaled seasonal ensemble predictions with the estimation of uncertainty. *J. Geophys. Res.* **116**: D08101, DOI: 10.1029/2010JD015284.

Palmer TN, Shukla J. 2000. Editorial to DSP/PROVOST special issue. *Q. J. R. Meteorol. Soc.* **126**: 1989–1990.

Palmer TN, Branković Č, Richardson DS. 2000. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Q. J. R. Meteorol. Soc.* **126**: 2013–2033, DOI: 10.1002/qj.49712656703.

Palmer TN, Doblas-Reyes J, Hagedorn R, Alessandri A, Gualdi S, Andersen U, Feddersen H, Cantelaube P, Terres J-M, Davey M, Graham R, Décluse P, Lazar A, Déqué M, Guérémy J-F, Díez E, Orfila B, Hoshen M, Morse AP, Keenlyside N, Latif M, Maionnave E, Rogel P, Marletto V, Thomson MC. 2004. Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER). *Bull. Am. Meteorol. Soc.* **85**(6): 853–872, DOI: 10.1175/BAMS-85-6-853.

Peng P, Kumar A, Van den Dool H, Barnston AG. 2002. An analysis of multimodel ensemble predictions for seasonal climate anomalies. *J. Geophys. Res.* **107**: 4710, DOI: 10.1029/2002JD002712.

Robertson AW, Lall U, Zebiak SE, Goddard L. 2004. Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Weather Rev.* **132**: 2732–2744, DOI: 10.1175/MWR2818.1.

Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, Van den Dool HM, Pan H-L, Moorthi S, Behringer D, Stokes D, Peña M, Lord S, White G, Ebisuzaki W, Peng P, Xie P. 2006. The NCEP climate forecast system. *J. Clim.* **19**: 3483–3517, DOI: 10.1175/JCLI3812.1.

- Stern W, Miyakoda K. 1995. Feasibility of seasonal forecasts inferred from multiple GCM simulations. *J. Clim.* **8**: 1071–1085, DOI: 10.1175/1520-0442(1995)008<1071:FOSFIF>2.0.CO;2.
- Sun J, Ahn J-B. 2011. A GCM-based forecasting model for the landfall of tropical cyclones in China. *Adv. Atmos. Sci.* **28**: 1049–1059, DOI: 10.1007/s00376-011-0122-8.
- Vasiliades L, Loukas A, Patsonas G. 2009. Evaluation of a statistical downscaling procedure for the estimation of climate change impacts on droughts. *Nat. Hazards Earth Syst. Sci.* **9**: 879–894, DOI: 10.5194/nhess-9-879-2009.
- Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press: New York, NY; 467.
- World Meteorological Organization (WMO). 2012. Guidelines on Ensemble Prediction Systems Forecasting. WMO No. 1091, Retrieved December 2, 2013. http://library.wmo.int/opac/index.php?lvl=notice_display&id=12962.
- Yu Z-P, Chu P-S, Schroeder T. 1997. Predictive skills of seasonal to annual rainfall variations in the U.S.-Affiliated Pacific Islands: canonical correlation analysis and multivariate principal component regression approaches. *J. Clim.* **10**: 2586–2599, DOI: 10.1175/1520-0442(1997)010<2586:PSOSTA>2.0.CO;2.
- Yun WT, Stefanova L, Krishnamurti TN. 2003. Improvement of the multimodel superensemble technique for seasonal forecasts. *J. Clim.* **16**: 3834–3840, DOI: 10.1175/1520-0442(2003)016<3834:IOTMST>2.0.CO;2.
- Yun WT, Stefanova L, Mitra AK, Vijaya Kumar TSV, Dewar W, Krishnamurti TN. 2005. A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. *Tellus* **57A**: 280–289, DOI: 10.1111/j.1600-0870.2005.00131.x.
- Zhao M, Hendon HH. 2009. Representation and prediction of the Indian Ocean dipole in the POAMA seasonal forecast model. *Q. J. R. Meteorol. Soc.* **135**: 337–352.