A new statistical correction strategy to improve long-term dynamical prediction

Joonlee Lee | Joong-Bae Ahn

Climate Prediction Laboratory, Division of Earth Environmental Systems, Atmospheric Sciences, Pusan National University, Busan, South Korea

Correspondence
Joong-Bae Ahn, Climate Prediction Laboratory, Division of Earth Environmental Systems, Atmospheric Sciences, Pusan National University, Gumjung-gu, Jangjeon-dong, Busan 609-735, South Korea.
Email: jbahm@pusan.ac.kr

In this study, a new statistical strategy to improve the long-term prediction skill of a numerical model was developed. This new strategy begins by finding the major principal time series (PTs) in the observations using the self-organizing map (SOM) method. Next, values at the model grid points that are highly correlated with the observational PTs for each ensemble member (EM) are combined to yield a modelled PT. Finally, the model prediction is corrected using the model PTs from the previous step. As the predictors for correction are objectively selected from among the signals found in model prediction, automatically considering their statistical correlation with predictands, the correction strategy is relatively free from the problem of selecting the proper predictor compared to conventional statistical correction methods. In addition, SOM shows a better performance in classifying non-linear complex patterns than conventional data analysis methods, while both SOM and conventional methods such as the empirical orthogonal function show a comparable performance when classifying linear patterns. The new strategy is applied to the 12-month-lead sea surface temperatures hindcasted by the Pusan National University coupled general circulation model. After correction using the new strategy, temporal correlation coefficients and the hit rate are increased while normalized root mean square errors and the false alarm rate are decreased for each season and each lead time. The correction becomes more effective as the lead time increases. In particular, this correction effect is large over the region where the prediction skill without correction is apparently low, which implies that the biases leading to poor prediction skills are effectively reduced by the new strategy. Additionally, the prediction skill is steadily improved for all lead times as the number of EMs is increased, whereas it reaches a plateau when the number of neurons in the output layer of the SOM method exceeds a certain threshold.

KEYWORDS
bias correction, coupled general circulation model, model output statistic, Self-organizing map method

1 | INTRODUCTION

The general circulation model (GCM) is an important tool for producing weather and climate information. In particular, the coupled GCM (CGCM) has become the ultimate tool for predicting changes and variations in climate (Meehl, 1995). The CGCM simulates the evolution of the earth system, which consists of the atmosphere, ocean, land surface, sea ice, and vegetation simultaneously, thereby determining future changes in subsystems such as the atmosphere and the ocean by interactions and feedbacks among the subsystems. Starting with Gates et al. (1985), the
CGCM has been used in many ways in recent years, in line with the rapid development of computing resources and an increased understanding of the interactions among the sub-systems (Behera et al., 2006; Cherchi et al., 2008). The CGCM is now widely used not only in seasonal prediction in meteorological centres such as the National Centres for Environmental Prediction (Saha et al., 2006) and the European Centre for Medium-Range Weather Forecasts (Van Oldenborgh et al., 2005), but also in the prediction of long-term climate changes through international projects such as the Coupled Model Intercomparison Project Phase 5 (Taylor et al., 2012).

However, even state-of-the-art CGCMs show biases in model results (Wang et al., 2008). These biases not only make it impossible to use the raw model data directly but also have deleterious effects on the predictability of the models (Kug et al., 2008). The model biases can be divided into two parts: mean and perturbation (Ahn et al., 2012). The bias in the mean part of a model can be corrected relatively easily by removing the model climatology from the model output (Kug et al., 2008; Jo and Ahn, 2015). For the perturbations, more sophisticated statistical (or dynamical) postprocessing is needed to reduce the bias. Previous studies have corrected the bias of the perturbations using different model output statistical methods depending on the variable, time, and region (Wilks, 2006; Kug et al., 2008; Ahn et al., 2012).

Statistical methods used to correct model biases are largely classified into linear and nonlinear methods. The linear methods include simple linear regression and multiple linear regression, both of which typically use the synoptic-scale variables or teleconnection indices as predictors (Wigley et al., 1990; Min et al., 2014). Other types of linear methods are empirical orthogonal functions (EOFs; Feddersen and Tompkins, 2011), the singular value decomposition analysis (Cherry, 1996; Kang et al., 2004) that utilizes the covariance of a model and its observations, and canonical correlation analysis (Barnett and Preisendorfer, 1987; Feddersen et al., 1999) that uses linear relations between a model and its observations. Nonlinear methods include genetic algorithms (Holland, 1975; Ahn and Lee, 2016), based on the process of biological evolution, and artificial neural networks (ANN; Nasseri et al., 2008), based on the learning processes of the human brain. However, for these conventional statistical correction methods, the process of selecting a predictor statistically correlated with a predictand is crucial because the improvement of the prediction skill depends on the predictor (Kim, 2003; Lee et al., 2015).

The objective of this study was to improve the dynamical prediction skill of dynamical models through a new correction strategy. We suggest this effective correction strategy for the purpose of removing biases in the perturbation part of numerical model results. The principle behind the strategy is to find meaningful signals hidden in the dynamical model prediction and to correct the prediction using these signals. Because the predictors for correction are objectively selected from among the signals found in model prediction, by automatically considering their statistical correlations with predictands, the correction strategy is relatively free from the problem of selecting the proper predictor when compared to the conventional statistical correction methods. Additionally, the self-organizing map (SOM) method in this study is used for finding major principal time series (PTs) to be utilized as a reference to detect the signals hidden in the dynamical model prediction. The SOM generally shows a better performance in classifying nonlinear complex patterns than conventional data analysis methods, while both SOM and conventional methods, such as EOF, show a comparable performance when classifying linear patterns (Liu and Weisberg, 2005; 2011; Liu et al., 2006; Chu et al., 2012). A detailed description of the method is presented in Section 2.2. The predictabilities achieved by analysing the results with and without the correction are compared in Section 3. Section 4 discusses and summarizes our results.

2.1 DATA AND METHODOLOGY

The model data used in this study are the retrospective forecast output from Pusan National University (PNU) CGCM, a participant model of the Asia-Pacific Economic Cooperation Asia-Pacific Climate Center multimodel ensemble long-range prediction system (Sun and Ahn, 2011; 2015; Kim and Ahn, 2015). Ten ensemble members (EMs) are produced with initial conditions for each day from 6 to 15 November using a time-lag method. This is an effective approach to forming an ensemble forecast that is known to efficiently reduce uncertainties resulting from initial conditions (Branković et al., 1990; Stensrud et al., 2000; Lu et al., 2007). The ensemble data have lead times of 12 months integrated at November of each year for the period of 1980–2014. The ensemble mean is obtained from the simple composite mean of the 10 EMs (Jo and Ahn, 2015; Ahn and Lee, 2016). The 35-year analysis period consists of four seasons: December–January–February (DJF, 1.5–3.5-month lead prediction), March–April–May (MAM, 4.5–6.5-month lead prediction), June–July–August (JJA, 7.5–9.5-month lead prediction), and September–October (SO, 10.5–11.5-month lead prediction). To ensure that the method suggested in this study can improve not only seasonal prediction but also long lead time forecast, lead times longer than the seasonal time scale were analysed. Among the predictive variables of the CGCM, sea surface temperature (SST) was chosen as the variable to be applied for the bias correction using SOM.
Hadley Centre Sea Surface Temperature (HadISST) (Rayner et al., 2003) data were used to verify the performance of the model for a period of 35 years (1980–2014). Anomalies were calculated based on the climatology for the period of 1980–2014.

2.2 Correction strategy

The correction strategy suggested in this study consists of three steps (Figure 1). The first step is to detect a principal time series from the observation (OPTs) using SOM. The second step is to find a signal similar to the OPTs in the model prediction. The final, third, step is to correct the model prediction using the signal found in the second step. A detailed description of each step is provided in the following section.

2.2.1 First step: Finding OPTs

The first step in the process is finding the major PTs from input data (e.g., reanalysis data) to be utilized as a reference to detect the signals hidden in the dynamical model prediction. In this study, SOM was used to identify major signals from the input data. SOM, a type of ANN trained by unsupervised learning (Kohonen, 1990), is already a well-known method (Liu and Weisberg, 2011). The SOM is generally an effective and useful statistical tool for clustering similar data patterns and projecting high dimensional data onto low-dimensional data (Liu and Weisberg, 2005; 2011). The SOM method can classify data features more effectively than conventional data analysis methods, such as EOF and principal component analysis (PCA), for various types of input data with complex patterns (Liu et al., 2006; Liu and Weisberg, 2011). For example, Liu and Weisberg (2005) and Liu et al. (2006; 2007) compared SOM methods with EOF and explained that SOM patterns were more accurate and intuitive than the EOF method. Additionally, Annas et al. (2007) and Astel et al. (2007) showed that SOM exhibits a better performance than PCA.

The SOM consists of input and output layers composed of neurons (Kohonen, 1990; Chu et al., 2012). The neurons of the input and output layers have input and weight vectors, respectively (Liu and Weisberg, 2011). Because each input data vector is compared with the weight vector of each neuron in the output layer, all input vectors are fully connected to the neurons in the output layer (Kohonen et al., 1996). For the first iteration, a random weight vector is used for initialization in the output layer, and then the weight vector...
nearest to the input data vector is detected using the Euclidean distance (ED) as in Equation (1):

\[ \text{ED}_j = \| P - W_j \| = \sqrt{\sum_{i=1}^{N} (P_i - W_{ij})^2}, \]

where \( N \) denotes the number of the total training period. \( P \) and \( W_j \) indicate the input data vector in the input layer and the weight vector of the \( j \)th neuron in the output layer, respectively. Among the neurons in the output layer, the neuron with the shortest ED is called the winner neuron. The weight vectors of neurons adjoining the winner neuron within a neighbourhood radius calculated by the Gaussian function are updated with the weight vector of the winner neuron by updating Equation (2) as follows

\[
W_j(t+1) = \begin{cases} 
W_j(t) + H(t)(P(t) - W_j(t)), & j \in R_j(t) \\
W_j(t) & \text{Otherwise,}
\end{cases}
\]

where \( t \) denotes the number of iterations and \( R_j(t) \) is the predefined neighbourhood around the \( j \)th neuron. \( C_C \) and \( C_j \) represent the coordinates of the winner neurons and the \( j \)th neuron in the output layer, respectively. \( H(t) \) denotes the Gaussian function, commonly used as a neighbourhood function (Chu et al., 2012; Yuan et al., 2015). \( B(t) \) and \( \sigma^2(t) \) are, respectively, a monotonically decreasing learning coefficient that determines the speed of the learning process \([0 < B(t) < 1]\) and the amplitude determining the width of the neighbourhood function. As the weight vector of not only the winner neuron but also the neighbouring neuron is updated by Equation (2), the final output produced by the SOM is a composite result of the winner neuron associated with the input neuron and the neurons adjacent to the winner neuron interacting with each other (Liu and Weisberg, 2011; Chu et al., 2012).

The SOM_PAK package is used in this study (Kohonen et al., 1996). For efficient computation, a simple and basic configuration of 3 × 3 neurons is chosen for the output layer. Sensitivity experiments are carried out to determine the effects of neuron size on the prediction skill—these are

![FIGURE 2](image_url) Temporal correlation map between sea surface temperature of HadISST and each OPT for DJF. Reanalysis data such as HadISST are simply employed as input data for SOM; it is not necessary to specifically use SST. The percentage at the top right of each panel indicates the fraction of grid points represented by this neuron. The shaded area shows temporal correlation coefficients that are significant at the 90% confidence level.
discussed in Section 3. For the purpose of correcting the model-predicted global SST, global reanalysis SST data (HadISST) are employed as input data for the SOM. Nine OPTs representing the nine major interannual time evolutions of global SSTs over all grid points of HadISST are obtained by the SOM in the first step, as Figure 1. To examine where OPTs show statistically significant relationships with different regions of the ocean, a correlation map between the OPTs and the HadISST observations in DJF is presented as an example (Figure 2). The correlation maps for OPTs that are adjacent in the output layer show similar patterns, whereas those that are far apart in the output layer display different patterns. In particular, OPTs located in the corners of the output layer appear to be related to different regions in the correlation map of SST. For instance, OPT3 in the upper right corner of the output layers shows a positive relationship in the equatorial western Pacific and northern Atlantic regions, and a negative relationship in the eastern Pacific regions. On the other hand, OPT7 in the lower left corner is positively related to interannual variations in SST over the equatorial eastern Pacific and the Indian Ocean, and negatively related to interannual SST variations in the western Pacific Ocean. OPT9 in the lower right corner displays a positive relationship around the Indian Ocean and the northern part of the equatorial region over the Atlantic Ocean. Intermediate neurons, which are affected by more than one corner, show mixed features of the corner neurons. For example, OPT8, located between OPT7 and OPT9, shows mixed features of OPT7 and OPT9. OPT8 shows a statistically significant relationship in the equatorial eastern Pacific and the equatorial northern Atlantic, which are dominant patterns of OPT7 and OPT9, respectively. This occurs because the SOM updates the winner neuron and the neighboring neurons together (Kohonen et al., 1996). The nine OPTs classified through the SOM are PTs that are correlated with the main ocean basins such as the Pacific Ocean, the Indian Ocean, and the Atlantic Ocean.

2.2.2 | Second step: Reconstructing the PTs

In the second step, potential signals, called newly reconstructed principal time series (NPTs), are found using the prediction of the dynamical model itself. The NPTs are generated by averaging the time series at grid points having good correlations with the OPTs that are higher than a certain threshold. Here, the threshold is defined as a correlation value significant at the 99% confidence level in hindcasts obtained from the model EMs. Therefore, NPTs are signals from the model prediction resembling the nine major time series of HadISST sorted by the SOM. Figure 3 shows averaged maps of temporal correlation coefficients (TCCs) for each season. The TCCs quantify the similarity between a given OPT and NPT, and the maps in Figure 3 also show regions where there is a statistically significant relationship between SST and a given PT. The regions with high TCC are slightly different in each season. In the case of the OPT, the TCCs of the DJF and MAM (JJA and SO) are relatively high in the equatorial and Indian Oceans (the South Pacific Ocean and the equatorial North Atlantic). The characteristics of the OPTs are well represented in the NPTs regardless of target seasons, although the regions of the NPT where TCCs are significant at the 95% confidence level are somewhat reduced compared to those of the OPT. The well-corrected values in certain areas are closely related to the NPTs in the new correction strategy. The regions with high TCC in the NPT's correlation map (Figure 3) correspond to the regions showing large correction effects because the NPT is used to modify the model prediction in the regions in the final step.

2.2.3 | Third step: Correcting the model prediction

In the third and final step, shown in Figure 1, the model prediction is corrected by combining the NPTs and the predictions of the model EMs for each grid. The criteria used to make suitable combinations of NPTs for each grid are based on TCCs between the model predictions and the HadISST (TCCmdl) and TCCs between the NPTs and the HadISST (TCCnpt). During the training period, the criterion used to select NPTs is that TCCnpt is larger than TCCmdl at those grid points. The average of the selected NPTs is used as the final corrected value for that grid point. If there are no selected NPTs (i.e., there are no grid points in which TCCnpt is larger than TCCmdl), the ensemble mean grid point is used without correction. The purpose of this process is to maintain the prediction skill of the model for grid points exhibiting good predictability because not every region can be corrected by NPTs (e.g., the white regions in Figure 3). The predicted value at each grid point satisfying the criterion is corrected using the combination of the NPTs. As the predictors for correction are objectively selected from among the NPTs, automatically considering the statistical correlation with predictands, the correction strategy is relatively free from the problem of choosing the proper predictor compared to the other statistical correction methods.

The corrected results are verified using cross-validation because the sampling period is not long enough to be subdivided into training and validation periods (Michaelsen, 1987; Barnston, 1994; Jo and Ahn, 2015). Because the commonly used leave-one-out cross-validation may result in overfitting due to the semibienniality of climates such as the Asian monsoon, this study applied leave-two-out cross-validation, excluding a prediction year and the following year, to assess the performance of the correction strategy.

3 | RESULTS

By comparing the prediction skills of the hindcasts with and without correction (referred to as corrected model result
(CMDL) and model result (MDL), respectively), the impact of the new correction strategy suggested in this study on forecast skill was estimated. Deterministic analyses such as TCC and normalized root mean square error (NRMSE; Jolliffe and Stephenson, 2003; World Meteorological Organization, 2006) were carried out for verification. TCC determined the linear correspondence between the observation (typically reanalysis data) and the model forecast with respect to time evolution (Wilks, 2006). The NRMSE, which normalizes the model RMSE with regard to the corresponding observed SD, provided the information about errors inherent in the forecast amplitude. Hit rates (HR) and false alarm rates (FAR), conventionally used for categorical deterministic analysis, have been also analysed (Wilks, 2006). The HR and FAR evaluate how well the model classifies the anomalies into the correct observation categories. In this

FIGURE 3 Spatial distribution of absolute temporal correlation coefficient between HadISST and OPTs (from a to d), and NPTs (from e to h) averaged over each season. Shading area shows temporal correlation coefficients significant at the 95% confidence level.
study, the MDL and CMDL anomalies were separated using their SD (σ) into three categories: larger than +0.43σ (+, above normal), smaller than −0.43σ (−, below normal), between −0.43σ and +0.43σ (0, normal).

Improvement in the prediction skill for SST was examined in terms of the spatial fields of the TCC (Figure 4). Shaded areas in Figure 4 indicate statistically significant TCC at the 90% confidence level, while dotted areas denote that the difference between the TCC skill of MDL and CMDL is significant at the 95% confidence level. TCC of the MDL is the highest in DJF, which is the season with the shortest lead time from the initial condition and decreases gradually as the lead time increases. Nevertheless, CMDL shows better results, indicating that the TCCs of CMDL are
higher in all seasons compared to those of MDL. In particular, the correction effect is larger in the region where the TCC between NPT and HadISST is high and the TCC between MDL and HadISST is low. This is because the correction is performed in the final step where statistically significant correlations exist between the NPTs and predictions of the dynamical model (Figure 1). For example, the prediction skill improved significantly in DJF and MAM in the Maritime Continent region, where the TCC of MDL was not statistically significant (Figure 4a and b) and the NPT was well correlated with HadISST (Figure 3e and f). Likewise, the prediction skill improved greatly in JJA and SO in the Indian Ocean and the Equatorial North Atlantic, where again, the TCC of MDL was not statistically significant (Figure 4c and d) and the NPT was well correlated with HadISST (Figure 3g and h).

Figure 5 shows the zonally averaged latitudinal variation of the TCCs over the global ocean for each EM, MDL, and CMDL for all seasons. MDL shows a better temporal correlation for EMs for all seasons, indicating that the ensemble mean approach is an effective way to improve the prediction skill (Hagedorn et al., 2005; Ahn and Lee, 2016). CMDL has a better correlation skill than MDL, and the latitudes with a large correction effect are different for each season. The TCCs of DJF and MAM (JJA and SO) show a large improvement around the equator (midlatitude of the Northern Hemisphere) because prediction skills are significantly enhanced in the Maritime Continent region (the Indian Ocean and the equatorial North Atlantic), as shown in Figure 4.

To evaluate various aspects of the improvement in prediction skill by the new correction strategy, seasonal (Figure 6) and monthly (Figure 7) mean SSTs of EM, MDL, and CMDL were each verified with respect to TCC, NRMSE, HR, and FAR. As lead time increased, the TCC and HR (NRMSE and FAR) decreased (increased), and the ensemble spread between the EMs increased (Figures 6 and 7). This is because the model uncertainty increases with increasing model integration, resulting in lower model predictability. These characteristics were also found in the CMDL results; however, the forecast skills of both the seasonal and monthly mean SST were evenly improved after correction. The CMDL results show an increase in TCC and HR and a decrease in NRMSE and FAR compared with those of MDL and EMs. In particular, for seasonal mean SST, CMDL indicates statistically confident prediction skill with temporal correlations reaching above the 95% confidence level at 5 months lead, whereas MDL retains prediction skill only out to 3.5 months lead. This implies that the corrected prediction can extend the statistically reasonable prediction for at least one season compared to the prediction by CGCM without correction. The improvement of prediction skill after correction is larger in SO (10.5–11.5 months

![Figure 5](image-url)  Zonally averaged (0°–360°E) temporal correlation coefficients of SST for the MDL, CMDL, and 10 ensemble members (EMs) during (a) DJF, (b) MAM, (c) JJA, and (d) SO. Lines with closed circle, open circle, and open square markers indicate EMs, MDL, and CMDL, respectively.
lead) than in DJF (1.5–3.5 months lead) as shown in Figures 6 and 7, indicating that a lower initial prediction skill leads to corresponding larger correction effects.

To examine how the correction effect depends on the prediction skill of the EMs, the TCC of the ensemble means was plotted against the TCC of the EMs in Figure 8 for both MDL (shown in circle) and CMDL (shown in square). This plot allowed us to compare the TCC yielded by the new ensemble method with that yielded by a simple composite method (Palmer et al., 2004; Jo and Ahn, 2015; Ahn and Lee, 2016). When the circle (MDL) fall above the reference line (1:1 line) in Figure 8, the TCC of the EMs is above zero; when they fall below the line, the TCC of the EMs is below zero. Figure 8 shows that an improvement in predictability is not expected with a conventional ensemble approach such as MDL when the prediction skills of EMs are poor, especially when they have a negative relationship with observations (TCC below 0). On the other hand, the prediction skill of CMDL is higher than that of MDL, regardless of the skills of the EMs, although the correction effects differ depending on the skill of EM. The prediction skill of CMDL is comparable to that of the MDL and better than those of the EMs over the region where the TCC of EMs is larger than approximately 0.7. As the TCCs of EMs decreases from 0.7, the correction effect gradually increases. The correction effect becomes great over the region where the skill of MDL is lower than the reference line and the correlation skill of the EMs is less than zero. The largest correction effect occurs when the correlation skill of EM is around −0.4 (Figure 8). This indicates that biases causing poor prediction skills are effectively reduced by the new correction strategy, even when EMs have negative correlation skills.

The improvement in prediction skill yielded by this correction strategy may depend on the size of the neurons in the output layer, which determines the number of OPTs in the SOM, and the number of EMs used to determine the similarity between OPT and NPT. Therefore, sensitivity tests are performed to investigate the effects of these two factors. Figure 9 shows the TCC at evaluating the change in prediction skill as a function of the number of neurons in the output layer and the number of EMs. The correlation skill improves as the numbers of both neurons (Figure 9a) and EMs (Figure 9b) increase, and this improvement depends slightly on the season, which is related to the different lead times associated with the different seasons. The enhancement of prediction skill is the
largest (smallest) in the SO (DJF) season, which has the longest (shortest) lead time, regardless of the number of neurons and EMs. Additionally, at all lead times, the prediction skill is steadily enhanced as the number of EMs increases. In contrast, the prediction skill remains stable once the number of neurons exceeds a certain number, as shown in Figure 9a. This number depends on the lead time (e.g., 2 × 2 and 4 × 4 neurons in the output layer are sufficient for improving the prediction skill in DJF and SO, respectively). These results show that the minimum number of neurons required to enhance prediction skill is greater at longer lead times.

4 | DISCUSSION AND SUMMARY
As the atmosphere is chaotic in nature and numerical models are hampered by important factors such as imperfection of numerical schemes and parameterization, the numerical model predictions include uncertainties and biases (Charney, 1951; Lorenz, 1963; IPCC, 2007). These uncertainties and biases increase with increasing model integration and render the signals predicted by the models less meaningful. This study aims to improve the forecast skill of a CGCM to correct the model prediction using meaningful signals from model prediction itself. The data used in this study are 35-year long 12-month lead PNU-CGCM hindcasts.

The correction strategy suggested in this study consists of three steps. In the first step, SOM is used on observational data to classify OPTs, which can be understood as the major patterns in the time series of global SST. The second step reconstructs NPTs that are similar to OPTs by averaging the time series of grid points showing higher TCC with OPTs than a given threshold, which in this case was chosen as a critical correlation value of a 99% confidence level. In the
third step, the model prediction is corrected using the combinations of NPTs from the previous step. The correction strategy is relatively free from the problem of selecting the proper predictor compared to the conventional statistical correction methods because the predictors for correction are objectively selected from among the signals found in the model prediction, automatically considering their statistical correlations with predictands. Additionally, SOM shows a better performance in classifying nonlinear complex patterns than the conventional data analysis method, while both SOM and conventional methods, such as the EOF, show a comparable performance when classifying linear patterns.

The TCCs of CMDL show improvement in all seasons and at all lead times compared with those of MDL. In particular, CMDL has a correlation skill above the 95% confidence level up to a 5-month lead time, whereas MDL has that only up to a 3-month lead time. The NRMSE of CMDL is also reduced in all seasons and at all lead times. HR and FAR, which are commonly used for a categorical deterministic analysis, corroborate these results. CMDL increases the HR and decreases the FAR, compared to MDL, for all seasons and lead times. The correction effect is large in the regions closely related to NPTs because NPTs are used to correct the model predictions in the final step. Additionally, the correction effect is large over the region where the MDL has poor prediction skill and the correlation skills of the EMs are less than zero, and is the largest over the region where the correlation skills of the EMs are around −0.4. This is because the biases that lead to lower prediction skills are effectively reduced by the new strategy. Conversely, models with higher prediction skills experience smaller improvements (i.e., when the model predictability is particularly good, only limited improvements in prediction skill can be obtained by the statistical method).

The improvement in prediction skill based on this correction strategy depends on the number of neurons in the output

---

**FIGURE 8** Scatter plot of the correlation skill of the ensemble mean among the ensemble members within each bin (y-axis) as a function of the ensemble member (x-axis). The bin size of the x-axis is 0.02, and only the average value for each bin range of the x-axis is marked. Open circle and square indicate the skill of MDL and CMDL, respectively.
layer and the number of EMs. The prediction skill continuously improves at all lead times as the number of EMs is increased, whereas the prediction skill reaches a plateau when the number of neurons exceeds a certain number. This number is large when lead time is longer and prediction skill is lower. However, regardless of these effects, CMDL performs better than MDL in all seasons.

Consequently, the new correction strategy suggested in this study is an effective way to reduce bias in the model and improve prediction skill, irrespective of season and lead time. In fact, the correction effect becomes larger as the lead time increases. The new strategy developed in this study can be utilized for various variables other than SST, the focus of this study. Additionally, any target region of interest can be corrected by the information embedded in other data, such as different kinds of reanalysis data, EMs, or the forecast itself. This strategy could be effectively used to improve dynamical prediction skill, because the predictors for correction are objectively selected from among the signals found in model prediction itself, automatically considering their statistical correlations with predictands.

ACKNOWLEDGMENTS
This work was carried out with the support of "Korea Meteorological Administration Research and Development Program under Grant KMI2018-01213" and "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01345205)" Rural Development Administration, Republic of Korea.

REFERENCES

FIGURE 9 Global mean TCC skill as a function of (a) the number of neurons and (b) the number of ensemble members for all four seasons. Open circle, square, triangle, cross markers denote the skill of DJF, MAM, JJA, and SO, respectively. The solid line shows the skill of MDL, and the dotted line indicates the skill of CMDL.