



RESEARCH ARTICLE

10.1002/2016JD025151

Key Points:

- A new multimodel ensemble (MME) method that uses a genetic algorithm (GA) is developed and applied for the prediction
- Three MME methods using GA (MME/GAs) are examined in comparison with a simple composite MME strategy
- The predictability of the MME/GAs shows a greater improvement than that of SCM MME strategy, particularly in higher-latitude land areas

Correspondence to:

J. Lee,
leejl@pusan.ac.kr

Citation:

Ahn, J.-B., and J. Lee (2016), A new multimodel ensemble method using nonlinear genetic algorithm: An application to boreal winter surface air temperature and precipitation prediction, *J. Geophys. Res. Atmos.*, 121, doi:10.1002/2016JD025151.

Received 27 MAR 2016

Accepted 26 JUL 2016

Accepted article online 29 JUL 2016

©2016. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

A new multimodel ensemble method using nonlinear genetic algorithm: An application to boreal winter surface air temperature and precipitation prediction

Joong-Bae Ahn¹ and Joonlee Lee¹
¹Division of Earth Environmental Systems, Atmospheric Sciences, Pusan National University, Busan, South Korea

Abstract A new multimodel ensemble (MME) method that uses a genetic algorithm (GA) is developed and applied to the prediction of winter surface air temperature (SAT) and precipitation. The GA based on the biological process of natural evolution is a nonlinear method which solves nonlinear optimization problems. Hindcast data of winter SAT and precipitation from the six coupled general circulation models participating in the seasonal MME prediction system of the Asia-Pacific Economic Conference Climate Center are used. Three MME methods using GA (MME/GAs) are examined in comparison with a simple composite MME strategy (MS0): MS1 which applies GA to single-model ensembles (SMEs), MS2 which applies GA to each ensemble member and then performs a simple composite method for MME, and MS3 which applies GA to both MME and SME. MS3 shows the highest predictability compared to MS0, MS1, and MS2 for both winter SAT and precipitation. These results indicate that biases of ensemble members of each model and model ensemble are more reduced with MS3 than with other MME/GAs and MS0. The predictability of the MME/GAs shows a greater improvement than that of MS0, particularly in higher-latitude land areas. The reason for the more improved increase of predictability over the land area, particularly in MS3, seems to be the fact that GA is more efficient in finding an optimum solution in a complex region where nonlinear physical properties are evident.

1. Introduction

The predictability of dynamic weather forecasts utilizing numerical models is limited by numerous factors such as the uncertainty of initial conditions, systematic errors of the model, and imperfections of the numerical schemes and parameterizations [Intergovernmental Panel on Climate Change (IPCC), 1996]. These factors result in inaccuracies of weather and climate predictions, to which various improvements have been suggested. The continuous efforts made to improve climate prediction have included the development of high-resolution models and the refinement of various physical processes (such as carbon feedback and atmospheric chemical reactions), in addition to various other approaches such as data assimilation and development of coupled general circulation model (CGCM) [IPCC, 2007]. The ensemble method has also been used recently to improve prediction probability by configuring numerous forecast members that are obtained by allowing perturbation in initial conditions [Houtekamer and Derome, 1995; Stensrud et al., 2000], or by altering the initial times of models [Lu et al., 2007], or by replacing physical processes [Stensrud et al., 1999]. Ensemble forecasting reduces the uncertainty of prediction resulting from errors in initial condition. The forecast is known to average out errors included in predictions obtained from different initial conditions [e.g., Kharin et al., 2001].

Recent recognition has acknowledged the effectiveness of the multimodel ensemble (MME), as this MME improves seasonal predictability in comparison with the single model (SM) by offsetting the systematic biases and errors of each climate prediction model. MME combines ensemble prediction data from various climate prediction models that have different dynamical and physical characteristics [Fraedrich and Smith, 1989; Krishnamurti et al., 1999]. MME is currently used across the whole spectrum of forecast ranges, including typhoon forecasts. Among the various linear methods used in MME [e.g., Kurihara et al., 1995; Palmer et al., 2004], Kalnay and Ham [1989], Fritsch et al. [2000], and Min et al. [2014] have shown that the prediction using the simple composite method (SCM), where the same weighting is given to each model, provides a better predictability than individual predictions. Krishnamurti et al. [2000] have also shown that the MME prediction using multiple linear regression (MLR) analysis has a higher predictability than the best SM prediction, since it removes the systematic errors of the models.

However, the conventional SCM is limited because it is useful only when all the SMs have similar and reasonable performances [Kug *et al.*, 2008b], and MLR is not an improvement over SCM because of its sampling problems, particularly when using hindcast data from a relatively short time period [Kharin and Zwiers, 2002; Peng *et al.*, 2002]. In addition, since MLR assumes a linear relationship between independent and dependent variables, a nonlinear relationship often existing among the variables cannot be explained by the method, even when the relationship between two variables is significant [Yuval and Hsieh, 2003]. Van den Dool and Rukhovets [1994] suggested that MME prediction can be improved by giving more weight to the member having relatively higher predictability, instead of assigning the same weight to all the members (if the predictability of each ensemble member is different). Yun *et al.* [2003, 2005] claimed that the use of the singular value decomposition and empirical orthogonal functions analysis methods affords the improved seasonal predictability compared to using SCM or SM when the MME strategy combines MLR.

The MME prediction strategy requires an applied nonlinear method because of the dynamical and physical nonlinear properties of the climate system. Yun and Krishnamurti [2002] and Park *et al.* [2005] used a nonlinear MME strategy to reflect the chaotic characteristic of the atmosphere and introduced a multilayer perceptron (MLP) of artificial neural network (ANN) for their strategy. However, MLP has the problem of stopping, even when the cost function reaches the local minima during the learning process [Choi *et al.*, 2000].

This study uses the genetic algorithm (GA) as an artificial intelligence (AI) technology which is capable of providing an efficient and flexible way to find an optimum solution in a complex searching space with a large nonlinear property. As first designed by Holland [1975], GA is a probabilistic method that expresses a series of processes for generating multiple descendants. It simulates evolution in an environment of “the survival of the fittest” to preserve genes and to ensure survival of the algorithm. When a new group is formed, the population observed as having high fitness in the previous group evolves into a new group using reproduction, crossover, and mutation of its genetic operators to accomplish an optimal solution that is then observed as having high fitness [Charbonneau and Knapp, 1995]. Coulibaly [2004] indicated that downscaling using GA is simpler and more efficient in the case of daily maximum and minimum temperatures than that using other statistical methods, and Nasseri *et al.* [2008] attempted to optimize a scenario for the prediction of precipitation using GA with ANN.

In this study, three MME experiments are designed to find a new MME method using GA and experimental results for the boreal winter prediction are compared to those of SCM, the conventional method commonly used in MME. Models participating in the Asia-Pacific Economic Conference Climate Center (APCC) multimodel seasonal prediction system are utilized for the analysis, and the target season and variables are the boreal winter and the surface air temperature (SAT) and precipitation, respectively. The MME method using GA is similar to the weighted composite method. However, the predictability can be different because GA finds the optimal weights by considering both linear and nonlinear relationships between the members. The experimental design is detailed in section 3.

2. Data and Method

2.1. Data

This study utilizes ensemble members of CGCMs produced from six institutions participating in the APCC program for MME long-range prediction. Table 1 shows the main characteristics of the CGCMs. All models have ensemble members to reduce the uncertainty in initial conditions, and hindcast sets of all models satisfy the requirements for the Seasonal Prediction Model Intercomparison Project/Historical Forecast Project (SMIP/HFP) and/or Coupled Model Intercomparison Project (CMIP).

Hindcast and forecast data of 1–3 month lead times from the initial condition of November are used. The seasonal mean SAT and precipitation from December, January, and February (DJF) are analyzed using the data of each model. A split period totaling 27 years from 1983 to 2005 and from 2008 to 2011 is used. The years 2006 and 2007 are omitted from the data because all models have generated hindcast data for the common period 1983–2005 and forecast data for the common period 2008–2011. The National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP/NCAR) reanalysis 2 data are used for the verification of 2 m temperature [Kanamitsu *et al.*, 2002]. For precipitation, monthly averaged data from the Global Precipitation Climatology Project are utilized [Huffman *et al.*, 1997].

Table 1. Description of Six CGCMs

Institute	AGCM (Resolution)	OGCM (Resolution)	Ensemble Member	Reference
PNU (Pusan National University)	CCM3 (T42 L18)	MOM3 (128 × 64 Gaussian grid L29)	5	<i>Sun and Ahn</i> [2011, 2014]
UH (University of Hawaii)	ECHAM4 (T31 L19)	UH Ocean (1° latitude × 2° longitude L2)	10	<i>Fu and Wang</i> [2004]
NCEP (National Centers for Environmental Prediction)	GFS (T62 L64)	MOM3 (1/3° latitude × 1° longitude L40)	15	<i>Saha et al.</i> [2006]
BMRC (Bureau of Meteorology Research Center)	BAM3.0d (T47 L17)	ACOM2 (72 × 144 grid points for the physics grid)	10	<i>Zhong et al.</i> [2005]
SNU (Seoul National University)	SNU (T42 L21)	MOM2.2 (1/3° latitude × 1° longitude L32)	6	<i>Kug et al.</i> [2008a]
APCC (Asia-Pacific Economic Conference Climate Center)	CAM3 (T85 L26)	POP1.3 (gx1v3)	5	

Observation and hindcast data are divided into mean (\bar{A}) and perturbation (A') using a perturbation method as follows.

$$A = \bar{A} + A'$$

Since the mean bias of the model can easily be removed by using observation and model climatologies [Kug et al., 2008c; Ahn et al., 2012], the MME strategy is applied only to A' , the perturbation part.

2.2. Method

2.2.1. Simple Composite Method (SCM)

SCM is a widely used and effective method for MME [e.g., Palmer et al., 2004; Wang et al., 2009; Min et al., 2014]. The SCM equation is as follows:

$$Y_t = \frac{1}{N} \sum_{i=1}^N A'_{i,t}$$

where $A'_{i,t}$ is the hindcast anomaly, t the time, and N the total number of input data. This method, which imposes the same weight to each model, is applied to each grid point. The previous studies demonstrated that some deterministic ensemble methods and probabilistic ensemble methods can have superior performance to SCM [e.g., Kug et al., 2008b; Suh et al., 2012; Oh and Suh, 2016]. Nevertheless, if all the input data have similar and reasonable performances, SCM generally outperforms other MME methods [e.g., Kug et al., 2008b; Min et al., 2014]. Min et al. [2014] have shown that SCM generally outperforms multiple regression-based weighted MME methods for all variables and seasons in APCC data. Therefore, in this study, we selected SCM, a simple and effective MME method, to compare with newly developed MME methods.

2.2.2. Genetic Algorithm (GA)

GA is a nonlinear optimization method based on the biological evolution to find a true value by mimicking the process of evolution in natural genetics [e.g., Holland, 1975]. A key concept of GA is the introduction of a chromosome that consists of genes in the process. A population is made of multiple chromosomes, and a generation is a process of evolving from one population to the next through the parent chromosome producing offspring chromosomes. A fitness function, which determines whether each chromosome is suitable to survive in the environment or not, is applied to this process. In GA, this fitness function sets the direction of evolution according to the user's configuration [Lee et al., 2006]. Operators used in the generation process of the GA are selection, crossover, mutation, and replacement. The selection, which is a fundamental operation in GA, is a process for selecting an outstanding chromosome preferentially. The crossover is a process of passing the gene from parent chromosome to offspring chromosomes. The mutation is a process of mutational genetic modification different from the simple genetic inheritance. The replacement is a process that replaces the population to the next generation. This operator determines whether all of the population or only the superior members are to be changed.

In this study, the GA is used for MME to find the optimal weight among the models and among the ensemble members of each model. The GA package used in this experiment is PIKAIA 1.2 [Charbonneau, 2002]. The MME results with GA are verified by using leave-one-out cross validation because the common period (27 years, 1983–2005 and 2008–2011) of each model's hindcast is too short to give a meaningful result [e.

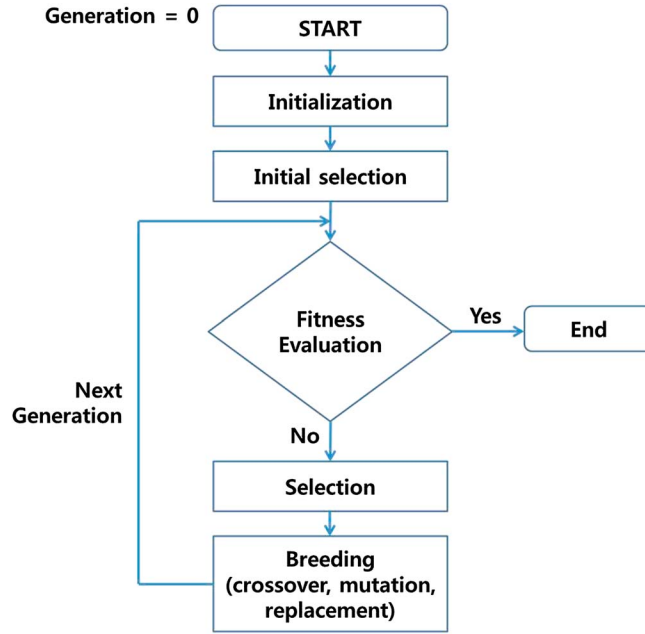


Figure 1. Flowchart of genetic algorithm process.

replacment until the optimal solution is attained. The fitness function used in this study is the minimum root-mean-square error (RMSE) between observation and hindcast anomalies. RMSE, which is a simple and effective fitness function, has been used by several previous studies [e.g., Lee *et al.*, 2006; You *et al.*, 2012]. RMSE is defined as follows:

$$RMSE = \frac{1}{T} \sum_{j=1}^T (k_j - O_j)^2$$

$$k_j = \sum_{i=1}^N (A'_{i,j} \times w_i)$$

where T and N are the total number of time and input data, respectively. $A'_{i,j}$ is the hindcast anomaly of input data, which is either each model or ensemble member of each model, and w_i is the weight per grid for each input data produced from GA. O_j is the observation, and k_j is the MME anomaly, which is multiplied by the weight (w_i) obtained from each model through GA.

The weight (w_i) for each model obtained from the GA is applied to the hindcast of target year (t) of each model as follows:

$$\bar{W} = \sum_{i=1}^N w_i$$

$$nw_i = \frac{w_i}{\bar{W}}$$

$$Y_t = \sum_{i=1}^N A'_{i,t} \times nw_i$$

where t is a target year and \bar{W} the sum of all the weights (w_i). nw_i , which has a value less than 1, is the final weight of each input data. Y_t is the final MME result for the target year to which the final weight is applied. The final weight (nw_i) is normalized with respect to total number of input data (N), and it represents the weight of each input data as a ratio.

g., Michaelson, 1987; Barnston, 1994; Jo and Ahn, 2014]. Therefore, the GA is applied to each model data for 26 years (except for the target year to forecast) to find the optimal weight, which is applied to obtain the final MME result of the target year. Figure 1 shows the schematic flowchart illustrating the process of the GA algorithm. First, the hindcast (26 years) of each model except the target year data is given as the input data of GA, and then an initial weight of each model is randomly generated. After the weight has been applied to the MME process, the result is assessed by a fitness function to check whether the optimal solution is obtained. If an optimal solution is not acquired, the algorithm proceeds further repeatedly to the next generation through operators such as crossover, mutation, and

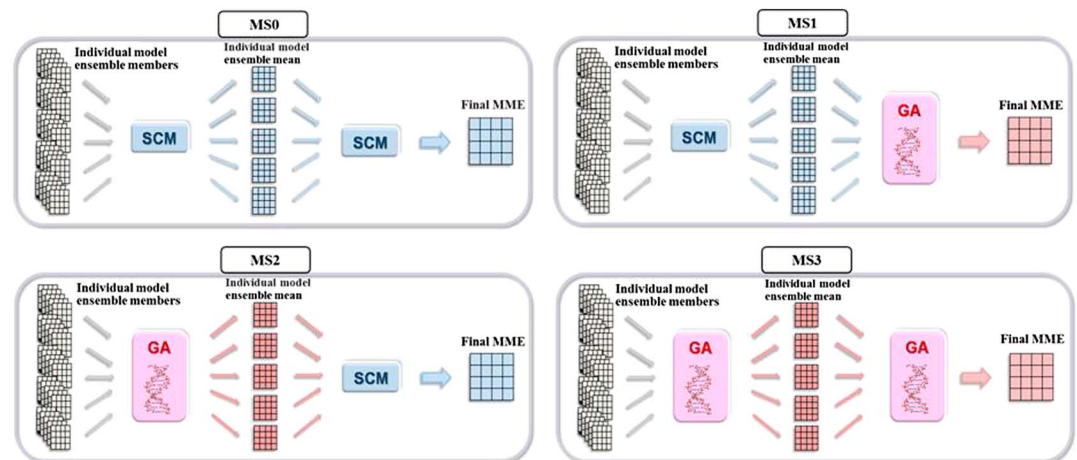


Figure 2. Description of the MS0, MS1, MS2, and MS3 methods.

3. Experimental Design

In this study, three MME experiments are designed to find the optimal MME method using GA and the experimental results for the boreal winter prediction are compared to those of SCM, the conventional method commonly used in MME (Figure 2). Applying a statistical method to the ensemble members of each model is referred to as a single-model ensemble (SME), and applying a statistical method to the SME results is referred to as an ESME.

The MME strategy 0 (MS0), a conventional method, applies SCM to both SME and ESME, and the result is used for comparison with other results obtained after GA application (Figure 2). The MME strategy 1 (MS1) is the one that obtains MME by applying SCM to the ensemble members of each model and then applies GA to that result. Because MS1 uses SCM, which gives the same weight to the input data, the predictability of the model can be reduced (particularly in the case of a large performance difference between the ensemble members). Moreover, there is a limit in finding the weight using the optimal fitness on each grid because the ensemble members of each model use the already-averaged result when applying GA. Thus, the performance of the final result can be inferior to that of the SME member which has the best performance. The second MME strategy (MS2) applies GA to SME and carries out SCM to ESME. This solves the problems of SME, but a limitation remains in improving the predictability because SCM is still included in the ESME process as in MS1. The third MME strategy (MS3) applies GA to both SME and ESME. This method not only benefits from an expansion of the searching space for finding the optimal weight but also compensates for the disadvantage of SCM which is lowering the predictability in the case of a large difference in the performances of the ensemble members. The results of SCM and GA methods used in the MME strategies are compared to each other for the prediction of the boreal winter SAT and precipitation over the period of 27 years (from 1983 to 2005 and from 2008 to 2011) on a global domain using ensemble members of six models developed by six institutions.

4. MME Strategy Performance Validation for New Techniques

The boreal winter (DJF) prediction of SAT and precipitation has been lead averaged for 1–3 months and analyzed from various perspectives because the model performance is assessed differently depending on analysis methods used to obtain skill scores [Hagedorn *et al.*, 2005]. The skill score is analyzed using the common deterministic analysis, as recommended by World Meteorological Organization, such as Temporal Correlation Coefficient (TCC), RMSE, Pattern Correlation Coefficient (PCC), and Normalized Standard Deviation (NSTD) and using the categorical deterministic analysis such as hit rate (HR) and false alarm rate (FAR) [Wilks, 1995]. For convenience, the six single models listed in Table 1 are arbitrarily referred to as SMs from SM1 to SM6 in order of model in Table 1.

Figure 3a shows TCCs of MS0, MS1, MS2, and MS3 compared with results from the six SMs. Overall, for SAT and precipitation, the methods that use the MME strategy show higher TCCs than SMs. The MME strategies using GA (hereafter, MME/GAs indicate MS1, MS2, and MS3) have higher TCCs than MS0, and among MME/GAs, MS3 produces the highest TCC. MS1 has a similar but slightly higher TCC than MS2.

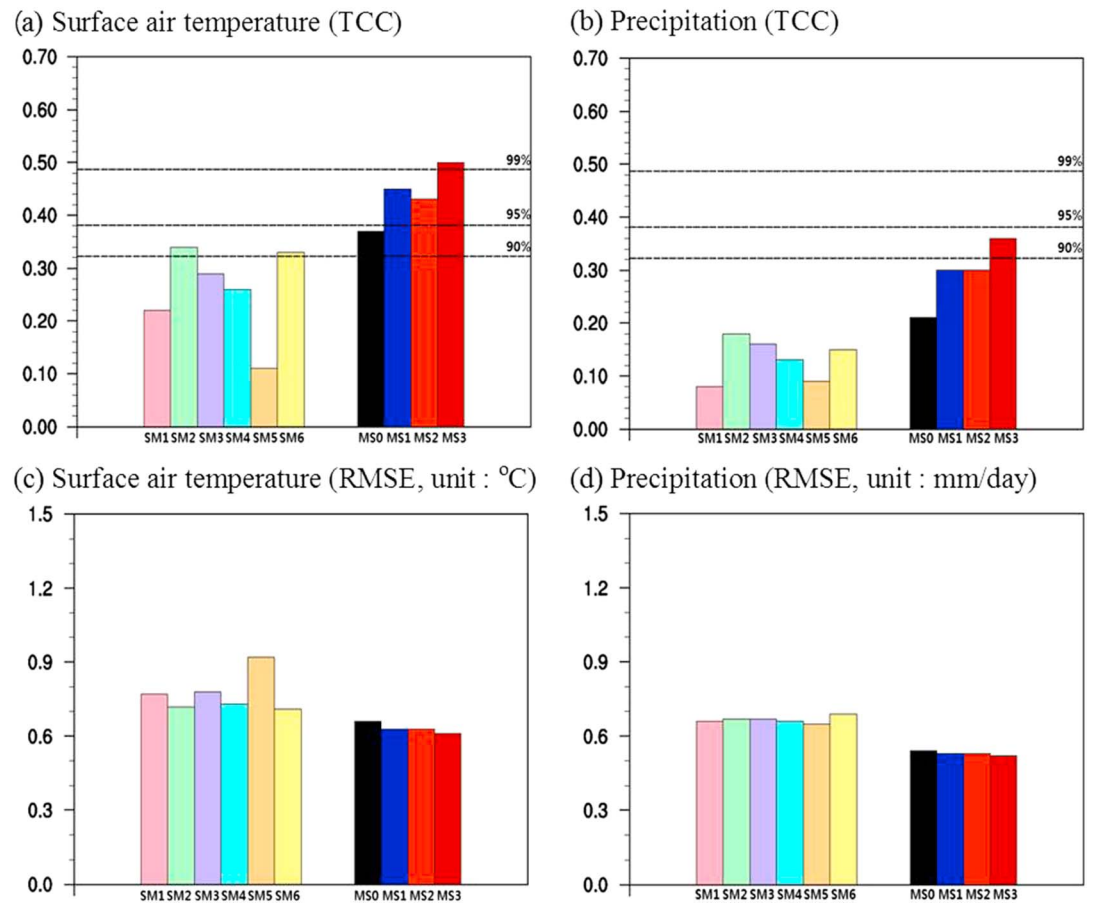


Figure 3. Global averages of (a and b) temporal correlation coefficient and (c and d) root-mean-square error of MMEs for SAT and precipitation compared with those of six single models (SMs).

Quantitatively, for SAT, MS1, MS2, and MS3 have TCCs of 0.45, 0.43, and 0.50, respectively, which are larger compared to those of SMs (about 0.26 on the average). All MME/GAs show significant correlations at more than 95% confidence level. MS3 shows a TCC increment of 0.16 compared to SM2, having the highest TCC among SMs, which has TCC exceeding more than 99% confidence level. MS0 has a TCC of 0.37, which is higher than any of SMs but lower than those of the other MME/GAs.

For precipitation (Figure 3b), SMs have a very low correlation coefficient of 0.13 on average, implying that the model predictability for precipitation is basically quite low compared to SAT. TCCs of MS0, MS1, MS2, and MS3 are 0.21, 0.30, 0.30, and 0.36, respectively. First of all, MS0 shows a higher value than any SMs. *Krishnamurti et al.* [1999, 2000], *Kharin and Zwiers* [2002], and *Min et al.* [2014] demonstrated the superior performance of SCM MME (MS0) prediction compared to the performance of each SM. *Hagedorn et al.* [2005] also showed that MME reduces the large amount of error compared to SM, in general. Our result corresponds with outputs of their studies. However, TCC of MS0 is lower than that of the other MSs using MME/GA. Compared to MS1 and MS2, even though the same GA is used, MS3 (which utilizes most of the characteristics of ensemble members and the weight by grid) shows a higher predictability for precipitation as well as SAT, with the confidence level of more than 90%. Unlike SAT, all of MME/GA results are below the 95% level of confidence for precipitation, although there are large increases in TCC compared to SMs.

Figures 3c and 3d show the RMSEs of SMs, MS0, MS1, MS2, and MS3. The results of RMSE are consistent with those of TCC. That is, the ensemble-averaged SM has the highest value, followed by MS0, MS2, MS1, and MS3. For SAT, the RMSEs of the SM average, MS0, MS1, MS2, and MS3, are 0.77°C, 0.66°C, 0.63°C, 0.63°C, and 0.61°C, respectively. The RMSE of MS3 is 0.16°C less than that of the SM average. The RMSEs for precipitation are 0.66 mm/d, 0.54 mm/d, 0.53 mm/d, 0.53 mm/d, and 0.52 mm/d for the SM average, MS0, MS1, MS2, and MS3, respectively. The RMSE of MS3 is about 0.14 mm/d less than that of SMs.

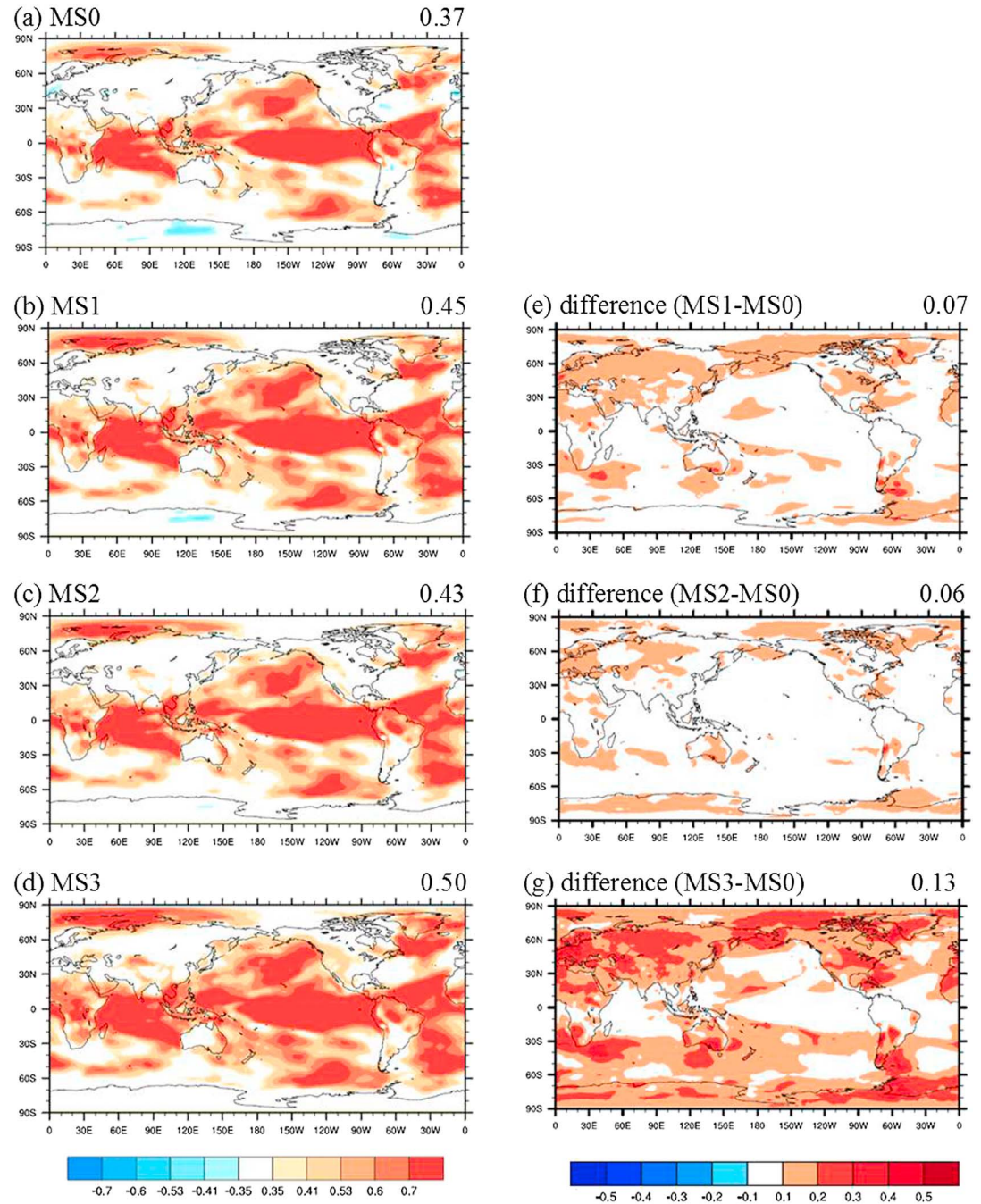


Figure 4. Temporal correlation coefficient of SAT over the global area for DJF. (a–d) TCC and (e–g) the difference between the MS0 and other methods. The colored area of TCC is significantly higher at the 90% confidence level. The numbers on the upper right of each figure represent the area-averaged TCC.

Figure 4 shows the TCC spatial distribution of SATs for MS0, MS1, MS2, and MS3 (Figures 4a–4d, respectively) and the spatial distribution of difference between MS0 and three GA methods (Figures 4e, 4f, and 4g). The numbers on the upper right of each figure represent the area-averaged TCC. MS0 has always a higher predictability than SMs but a lower predictability than MME/GAs. The predictability of MS1, MS2, and MS3 shows a greater improvement than that of MS0, increasing by 18.9%, 16.2%, and 35.1%, respectively. MS3 shows the greatest improvement in predictability, particularly in higher-latitude land areas where TCC is relatively low (Figure 4a). The predictability over the land area is improved to a greater extent because GA is more efficient in finding an optimum solution in a complex region where nonlinear physical properties are more possible.

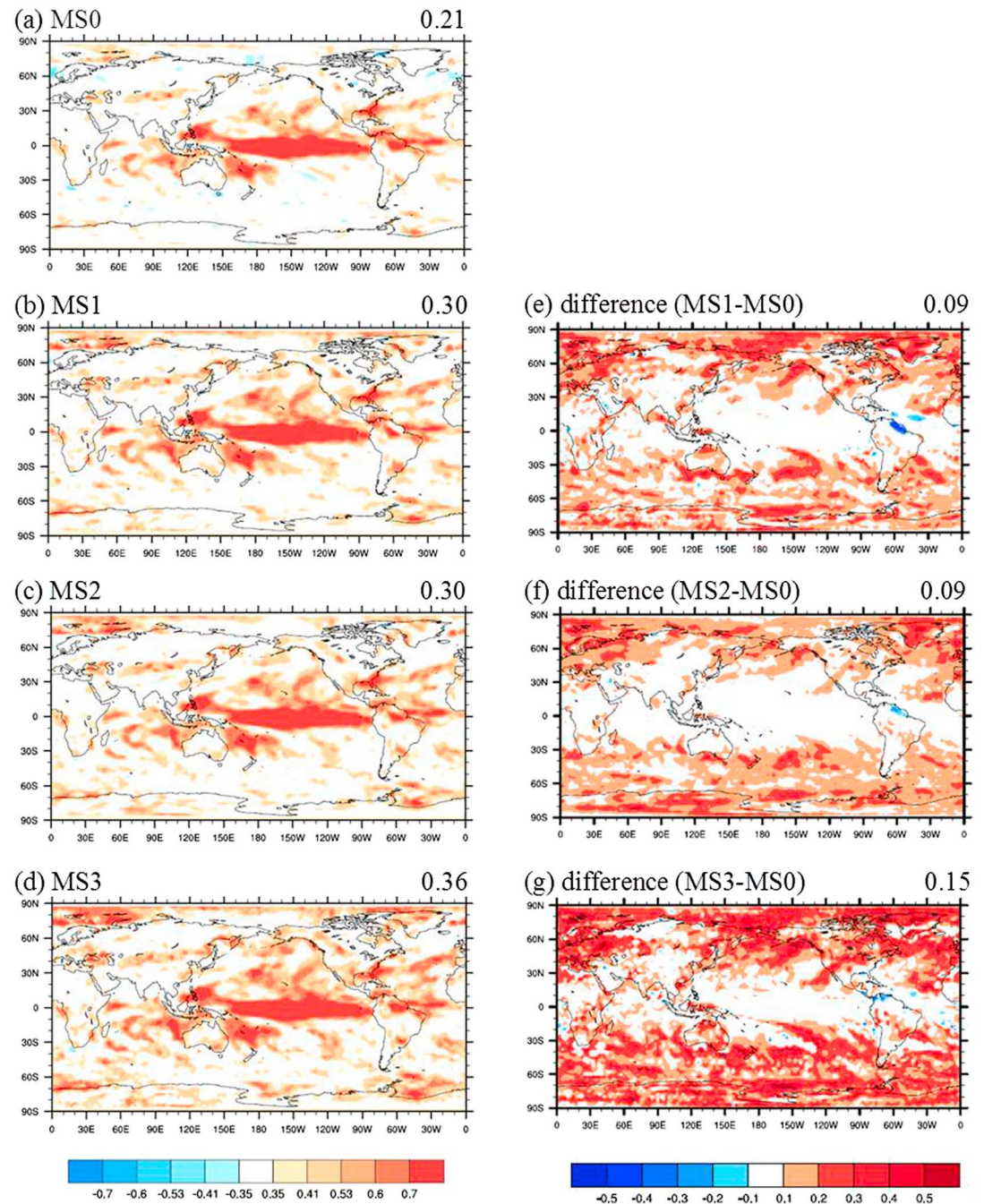


Figure 5. Temporal correlation coefficient of precipitation over the global area for DJF. (a–d) TCC and (e–g) the difference between the MS0 and other methods. The colored area of TCC is significantly higher at the 90% confidence level. The numbers on the upper right of each figure represent the area-averaged TCC.

This is because the GA method generates the optimal solution for a given problem using nonlinear functions such as inheritance, mutation, and selection [e.g., *Nasseri et al.*, 2008; *Tabassum and Mathew*, 2014].

Precipitation, the end product of atmospheric physical moisture processes, is a highly nonlinear property, and thus, it has a lower predictability than other atmospheric variables such as SAT. With linear SCM methods, precipitation is generally well predicted only around the equator (Figure 5). However, GA, which is a nonlinear method, shows a great improvement at high latitudes where the predictability is low (as that of SAT). For precipitation, the predictabilities of MS1, MS2, and MS3 show greater improvement than MS0 does,

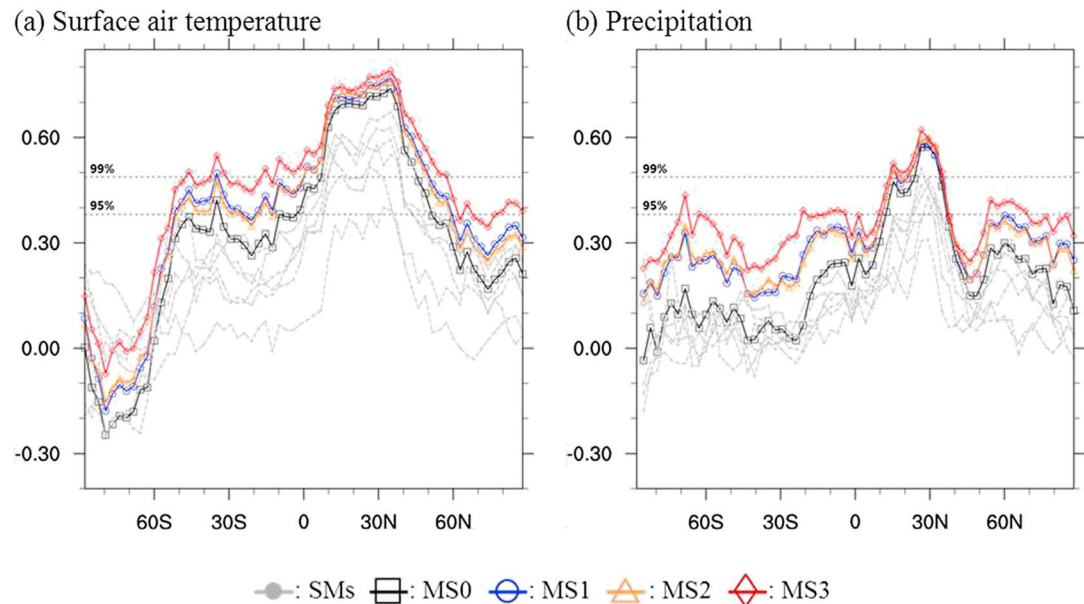


Figure 6. Zonally averaged (0° – 360° E) temporal correlation coefficients for the MS0, MS1, MS2, MS3, and six single models (SMs) for SAT and precipitation at each latitude.

increasing from 0.21 to 0.30, 0.30, and 0.36 (corresponding to 42.8%, 42.8%, and 71.4% increases), respectively. In particular, MS3 also shows the greatest improvement in predictability for precipitation among all three GA methods.

The zonally averaged (0° – 360° E) TCC between observation and the MSs (MS0 ~ 3), and six single models (SMs) for the SAT and precipitation with respect to latitude, are drawn in Figure 6. For both SAT and precipitation, as seen in Figures 3, 5, and 6, TCCs of MS0 are higher than those of SMs [Hagedorn *et al.*, 2005] and TCCs of MME/GAs are generally higher than those of MS0 at all latitudes. Among the MME/GAs, MS3 shows the highest TCC.

For SAT (Figure 6a), TCCs of MS0 are higher than those of SMs but lower than those of MME/GAs. MS3 has the best performance among the MME/GAs, followed by MS1 and MS2. TCCs are higher around the equator and lower in the higher latitudes. Particularly, to the south of around 60° S, TCCs of the MME/GAs even have opposite sign. This indicates that the improvement of predictability using the MME/GA methods can be limited by the performance of individual models. That is, the predictabilities of the MME/GA methods are limited where SMs exhibit poor performances. Kug *et al.* [2008b] and Min *et al.* [2014] also mentioned that the performances of the MME methods depend on those of the individual model.

Comparing Figure 6b with Figure 6a, TCCs of precipitation are lower than those of SAT, in general, because of the lower performance of SMs. TCCs of MS0 are higher than those of SMs, which corresponds with previous study finding. MME/GAs have higher predictabilities than SMs and MS0 for both variables, and especially, MS3 shows the most improved performance among MME/GAs throughout the latitudes. Unlike the case of SAT, the TCCs of MSs for precipitation are higher than SMs. This is because although the TCCs of each SM for precipitation are relatively low compared to SAT, they are not as extremely poor as SAT in the latitudes south of 60° S.

From this analysis, we find that the improvement in predictability relies on the performance of each individual SM. The improvement in the predictability of MMEs is large where SMs show relatively poor performances but small where SMs exhibit good performances. However, where the predictabilities of SMs are extremely poor, as in Figure 6a (south of 60° S), a large improvement of predictability utilizing MME/GAs was not expected. In general, relationship between the observation and model output obtained from the training period is applied to the predictive period in statistical method under a stationarity assumption. Such assumption implicitly used for the nonlinear as well as linear MME methods, which is a fundamental limitation of statistical method, may be one of the possible causes for the abnormal results sometimes

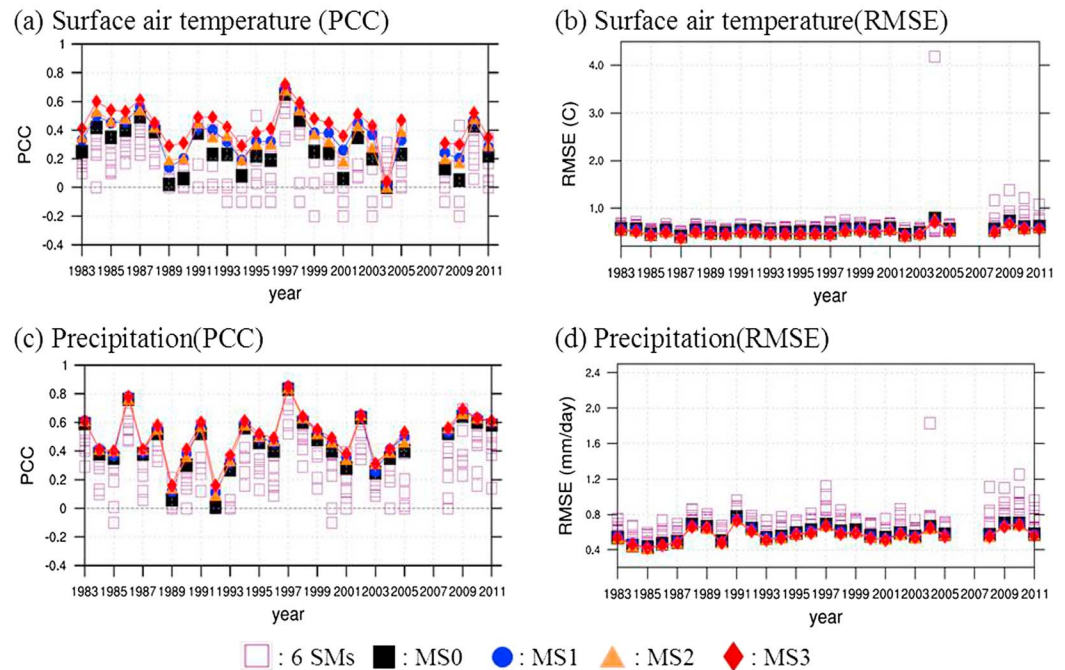


Figure 7. Time series of anomaly pattern correlation coefficients and root-mean-square error (RMSE) for the MS0, MS1, MS2, MS3, and six single models (SMs) for the global mean SAT and precipitation.

obtained in the high latitudes or specific years (e.g., 2004 in Figure 7). It is particularly plausible when one of the ensemble members has abnormally poor predictive skill at specific regions and periods.

Figure 7 presents the PCCs and RMSEs of SMs, MS0, MS1, MS2, and MS3 in a time series. The noncolored squares represent SMs, colored squares MS0, and circle, triangle, and diamond denote MS1, MS2, and MS3, respectively. For SAT, the MME/GAs show a better pattern correlation than SMs and MS0 in most years, and MS3 in particular shows the highest values. The RMSEs of MME/GAs are lower than those of SMs and MS0 as shown in Figure 7b. However, the PCCs of SAT for the MME methods (MS0, MS1, MS2, and MS3) in 2004 are even lower than those of SMs because one model generates a result with poor performance, as shown in Figure 7b. These results reveal that the performances of the MME methods are influenced by those of input data. Except for 2004, the overall predictabilities of the MME methods are better than those of the SMs. This is evident in Figure 8, where the total period mean of the PCC is shown along with RMSE. The decrease of RMSE and the increase of the averaged PCC are evident in MME/GAs, and MS3 shows the best predictability for SAT. MS0 shows lower RMSE than SMs. The MME/GAs also have higher spatial pattern correlations and lower RMSEs for precipitations than do the SMs and MS0. The PCC-RMSE diagram also shows that MME/GAs have a higher predictability than the SMs do. The PCCs in precipitation are generally higher than those of SAT in Figures 7 and 8 because the variance of global precipitation anomaly is generally concentrated at the equator and the overall characteristic of the anomaly spatial distribution of precipitation is more easily captured by models compared to SAT. However, it should be noted that the temporal variance of precipitation is lower than that of SAT (e.g., Figure 3).

Figure 9 presents Taylor diagrams for SMs and MME strategies which is drawn in order to investigate variations in SAT and precipitation. The Taylor diagram is a tool that compares similarities of patterns between the model and the observation by obtaining nondimensional parameters of NSTD and TCC. For SAT, results of MS0 and MME/GAs have significant correlations at the 95% confidence level. MS3 shows a significant correlation at more than 99% confidence level. The variation in MME strategy generally decreases in comparison with SM results. Furthermore, among MME strategies, GA methods show variations that are more similar to observations than variations of MS0. For precipitation, MS0 has a higher correlation with the observation than SMs, and MME/GAs show a higher correlation than MS0. According to previous PCC-RMSE and Taylor diagrams, the MME/GAs construct spatial and temporal patterns of SAT and precipitation better than SMs, and among MME/GAs, MS3 shows the highest predictability. MS1 and MS2 are

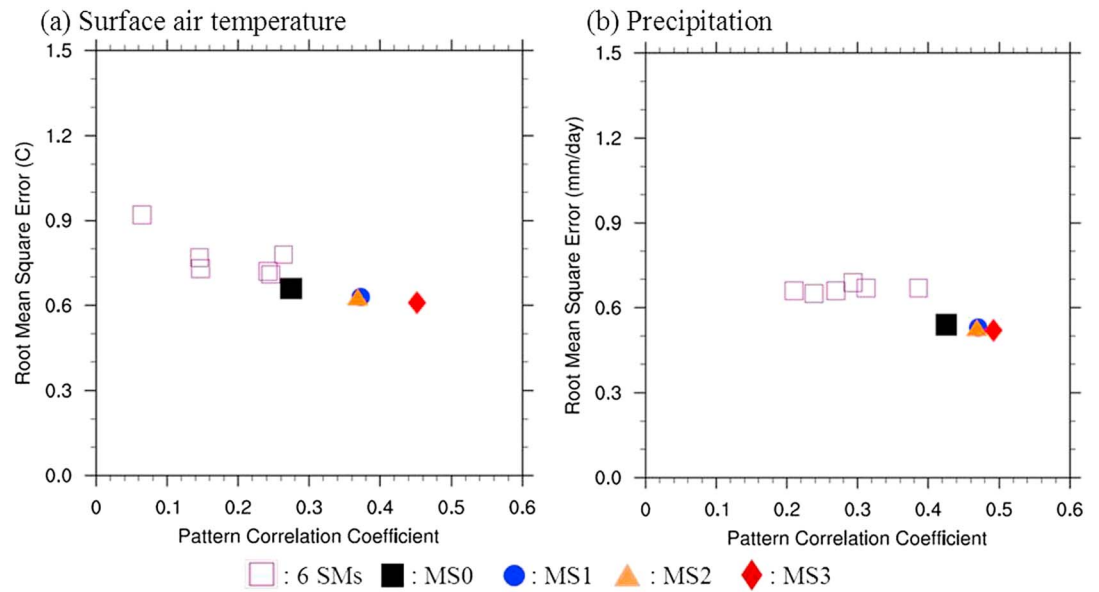


Figure 8. Anomalous PCC-RMSE diagram of MS0, MS1, MS2, MS3, and six single models (SMs) for the global mean (a) SAT and (b) precipitation.

similar to each other in terms of the predictability, but MS1 has a better output for the TCC of SAT and MS2 for the PCC of precipitation.

Figure 10 shows a diagram of HR and FAR, which exhibits the categorical predictability for boreal winter SAT and precipitation anomalies. HR and FAR are obtained by classifying the given values into three categories: larger than $+0.53\sigma$ (+), smaller than -0.53σ (-), and between $+0.53\sigma$ and -0.53σ (0) where, σ means the standard deviation of each MME result [e.g., Wilks, 1995; Jo and Ahn, 2014]. In Figure 10, the upper left corner of the chart indicates a perfect forecast and the reference line (diagonal line) a zero-skill forecast. MS1, MS2, and MS3 show improved skills overall compared to SMs and MS0 as their positions move toward the upper left corner of the chart by increasing HR and decreasing FAR for both SAT and precipitation. Again, MS3 provides the greatest skill among the MME/GAs, while MS2 and MS1 show no significant differences.

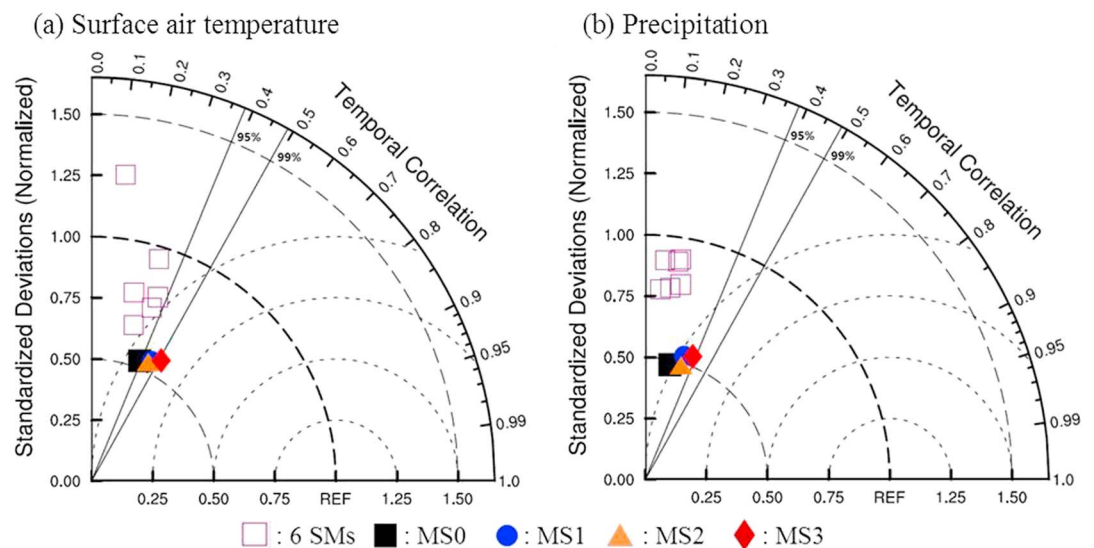


Figure 9. Taylor diagram of MS0, MS1, MS2, MS3, and six single models (SMs) for the global mean (a) SAT and (b) precipitation.

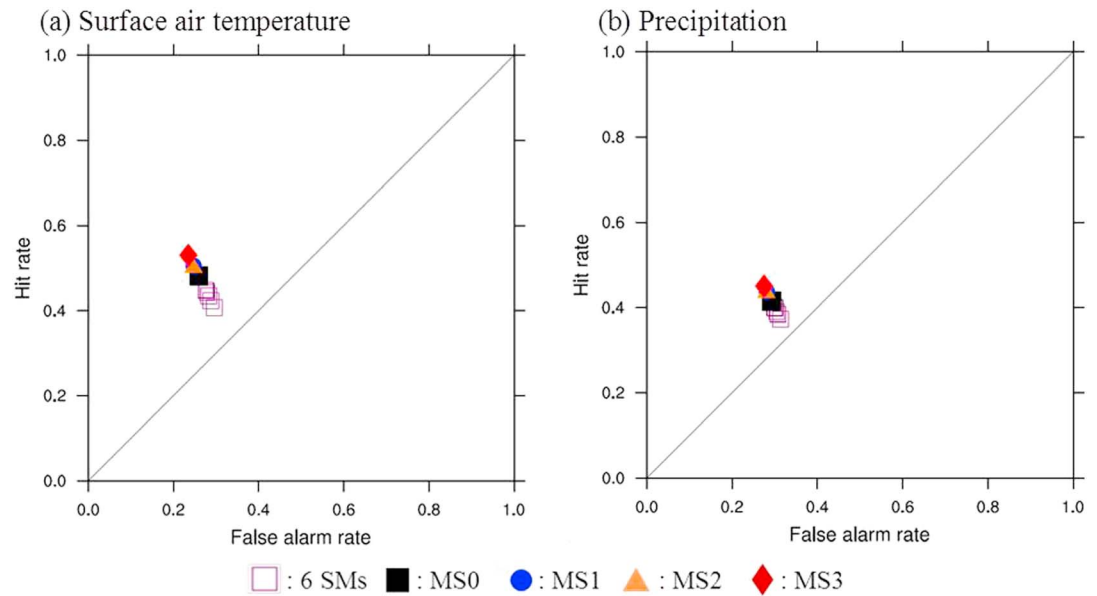


Figure 10. Hit rate (HR) and false alarm rate (FAR) diagrams of MS0, MS1, MS2, MS3, and six single models (SMs) for the global mean (a) SAT and (b) precipitation.

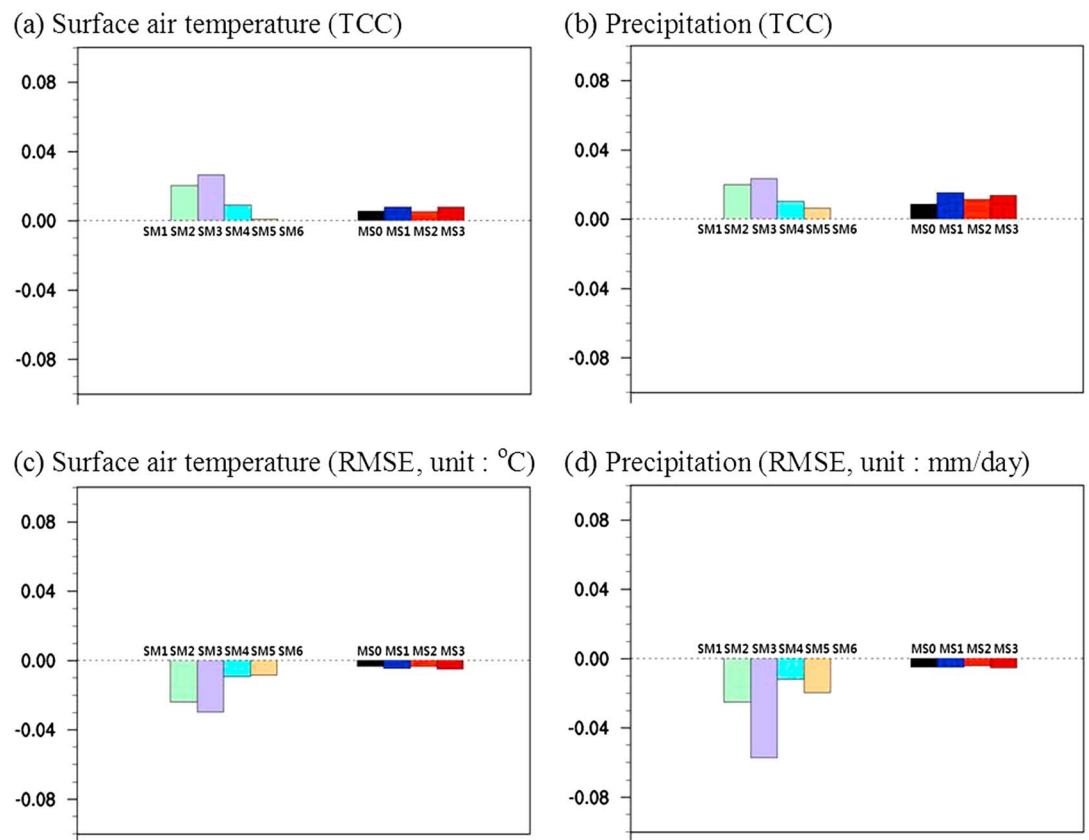


Figure 11. Global averaged differences between the maximum and minimum ensemble number experiments (a and b) temporal correlation coefficient and (c and d) root-mean-square error for SAT and precipitation comparing MMEs with those of the six single models (SMs).

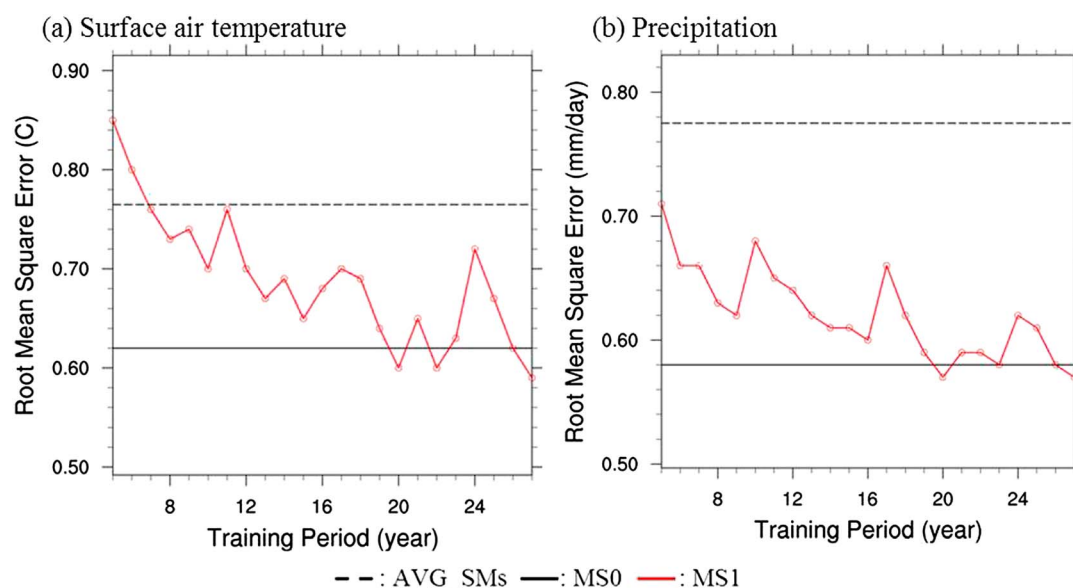


Figure 12. Root-mean-square error (RMSE) of MS1 in DJF of 2011 as a function of training period for global mean (a) SAT and (b) precipitation. The black and dashed lines denote the RMSEs of MS0 and averaged value of six single models (SMs), respectively.

Since the predictabilities of the MME methods can be changed by the number of ensemble members and training periods [e.g., *Oh and Suh, 2016*], sensitive tests for the number and period are conducted on MME/GAs. To examine the changes in performance in relation to the number of ensemble members, first, the MME methods are applied additionally to five-ensemble member, which is the minimum member among the models. Since the whole number of ensemble members for the two models (SM1 and SM6) are only 5, additional experiment for sensitivity test for the models is not performed. Figure 11 presents the differences of TCC and RMSE between the ensemble experiments with maximum (all ensemble members used) and minimum (five ensemble members used) ensemble members for SAT and precipitation. TCCs (RMSEs) of SMs increase (decrease) as the number of ensemble members increases for both variables. The predictability of each MME method is also changed as the number of ensemble changes. The results show that the changes of TCCs and RMSEs are generally smaller than those of SMs. This means that the MME methods are less dependent on and sensitive to the number of ensemble members compared to SMs.

To test the sensitivity of the MME methods in relation to the training period, RMSEs for boreal winter (DJF) of 2011 are examined by changing the training period from 5 to 27 years (Figure 12). The dotted and solid lines are the average of SMs (AVG_SMs) and MS0, respectively. Both SAT and precipitation show a decrease of RMSE as the training period increases. For SAT, MS1 exhibits better performance than AVG_SMs do if the training period is longer than 8 years. Precipitation shows better performance than AVG_SMs in all training periods, indicating that MME/GAs have superior performance in precipitation to SMs regardless of the training periods. In addition, RMSEs of MME/GAs are lower compared to MS0 when training periods are longer than 26 years.

5. Discussion and Conclusion

MME/GAs, which are nonlinear MME strategies capable of reflecting chaotic characteristics of the atmosphere, are developed as new MME methods to improve the seasonal predictability. The success of the proposed MME strategy is attributable to GA's AI technology, which determines the optimum solution flexibly and efficiently in a complex searching space with a large nonlinear property.

The winter SAT and precipitation predicted from ensemble members of six CGCMs participating in the APCC long-range MME prediction system are used. To optimize the MME strategy, the following three MME/GAs strategies are analyzed in comparison with the simple SCM MME strategy (MS0): MS1 which applies GA to SMEs, MS2 which applies GA to each ensemble member and then performs SCM for MME, and MS3 which applies GA to both MME and SME.

Compared to SMs and MS0, our results demonstrate that MME/GAs have a higher TCC and a lower RMSE for both SAT and precipitation. That is, for temperature, TCC of MS1, MS2, and MS3 shows an improvement than that of MS0, increasing from 0.37 of MS0 to 0.45, 0.43, and 0.50 (corresponding to 18.9%, 16.2%, and 35.1% increases), respectively. TCC for precipitation of MS1, MS2, and MS3 also increases from 0.21 to 0.30, 0.30, and 0.36 (corresponding to 42.8%, 42.8%, and 71.4% increases compared to that of MS0), respectively. PCC-RMSE and Taylor diagram analyses also corroborate these results. In addition, the MME/GAs show the superior forecast accuracy for both SAT and precipitation even in the categorical deterministic analysis, as the increase of HR and the decrease of FAR are noted globally. This indicates that MME/GAs improve the predictability by offsetting the bias of the poorly performing model with that of the skillful model [Hagedorn *et al.*, 2005]. In addition, MS3, which uses GA for both SME and MME, exhibits the greatest predictability for both SAT and precipitation compared to MS1 and MS2. The improvement in the predictabilities of MME/GAs is large where SMs exhibit relatively poor performances but small where SMs show good performances. However, a large improvement of predictabilities utilizing both MME/GAs and MS0 was not expected where the predictabilities of SMs are extremely poor as in Figure 6a (south of 60°S). It demonstrates that if the performances of individual models are extremely poor, the improvement of predictability using the MME methods can be limited.

TCCs (RMSEs) of SMs increase (decrease) as the number of ensemble member increases for both variables. The predictability of each MME method is also changed as the number of ensemble changes. However, the MME/GAs have superior performance and are less dependent on and sensitive to the number of ensemble members compared to SMs. Both SAT and precipitation exhibit a decrease of RMSE as the training period increases. For SAT, RMSEs of MS1 are lower compared to SMs if the training period is longer than 8 years. Precipitation shows better performance than SMs regardless of the training period. In addition, MME/GAs have superior performance to MS0 when training periods are longer than 26 years.

Our results show that MME/GAs can improve the predictability compared to SMs and MS0. This improvement is also dependent on how the GA method is applied. MS3, which uses the GA method to each model's ensemble members and each model, provides the highest predictability, compared to MS1 and MS2 (including SCM). MS2 and MS1, however, show no statistically significant difference. These results imply that biases of ensemble members of each model and model ensemble are reduced more with MS3 than with other MME/GAs.

The predictability of MME/GA methods shows a greater improvement than that of MS0, particularly in higher-latitude land areas. The reason for the more improved increase of predictability over the land area, particularly in MS3, seems to be the fact that GA is more efficient in finding an optimum solution in a complex region where nonlinear physical properties are more possible. Although MS3 has been applied to boreal winter SAT and precipitation in this study, the method is expected to be used effectively for any variables and seasons.

Acknowledgments

This work was carried out with the support of the Rural Development Administration Cooperative Research Program for Agriculture Science and Technology Development under grant project PJ009953 and the Korea Meteorological Administration Research and Development Program under grant KMIPA 2015-2081, Republic of Korea. The APCC model data were obtained from http://cis.apcc21.org/opendap/INDV_MODEL/3-MON/.

References

- Ahn, J. B., J. L. Lee, and E. S. Im (2012), The reproducibility of surface air temperature over South Korea using dynamical downscaling and statistical correction, *J. Meteorol. Soc. Jpn.*, *90*(4), 493–507.
- Barnston, A. G. (1994), Linear statistical short-term climate predictive skill in the Northern Hemisphere, *J. Clim.*, *7*, 1513–1564.
- Charbonneau, P. (2002), An introduction to genetic algorithms for numerical optimization, NCAR Tech. Note TN-450 IA, 74pp.
- Charbonneau, P., and B. Knapp (1995), A user's guide to PIKAIA 1.0, NCAR Tech. Note 418+IA, 121pp.
- Choi, S. S., H. K. Koo, and Y. K. Kim (2000), Predicting stock prices using book values and earnings-per-share based on linear regression model and neural network model, *Kor. J. Financ. Man.*, *17*, 161–180.
- Coulibaly, P. (2004), Downscaling daily extreme temperatures with genetic programming, *Geophys. Res. Lett.*, *31*, L16203, doi:10.1029/2004GL020075.
- Fraedrich, K., and N. R. Smith (1989), Combining predictive schemes in long-range forecasting, *J. Clim.*, *2*, 291–294.
- Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislocky (2000), Model consensus, *Weather Forecasting*, *15*, 571–582.
- Fu, X., and B. Wang (2004), The boreal-summer intraseasonal oscillations simulated in a hybrid coupled atmosphere-ocean model, *Mon. Weather Rev.*, *132*, 2628–2649.
- Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept, *Tellus A*, *57*, 219–233.
- Holland, J. H. (1975), *Adaption in Natural and Artificial Systems*, 228 pp., Univ. Michigan Press, Ann Arbor.
- Houtekamer, P. L., and J. Derome (1995), Methods for ensemble prediction, *Mon. Weather Rev.*, *123*(7), 2181–2196.
- Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider (1997), The Global Precipitation Climatology Project (GPCP) combined precipitation dataset, *Bull. Am. Meteorol. Soc.*, *78*, 5–20.
- Intergovernmental Panel on Climate Change (IPCC) (1996), *Climate Change 1995—The Science of Climate Change, Contribution of Working Group I to the Second Assessment Report of the IPCC*, Cambridge Univ. Press, New York.

- Intergovernmental Panel on Climate Change (IPCC) (2007), *Climate Change 2007—The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 793–795, Cambridge Univ. Press, Cambridge, U. K., and New York.
- Jo, S., and J. B. Ahn (2014), Improvement of CGCM prediction for wet season precipitation over Maritime Continent using a bias correction method, *Int. J. Climatol.*, *35*(13), 3721–3732.
- Kalnay, E., and M. Ham (1989), Forecasting forecast skill in the Southern Hemisphere, Preprints of the 3rd International Conference on Southern Hemisphere Meteorology and Oceanography, Buenos Aires, 1989, 24–27pp.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP-DEO AMIP-II Reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, *83*, 1631–1643.
- Kharin, V. V., and F. W. Zwiers (2002), Climate prediction with multimodel ensembles, *J. Clim.*, *15*, 793–799.
- Kharin, V. V., F. W. Zwiers, and N. Gagnon (2001), Skill of seasonal hindcasts as a function of the ensemble size, *Clim. Dyn.*, *17*, 835–843.
- Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiocchi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran (1999), Improved weather and seasonal climate forecasts from multi model superensemble, *Science*, *285*, 1548–1550.
- Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. E. LaRow, D. R. Bachiocchi, C. E. Williford, S. Gadgil, and S. Surendran (2000), Multi-model ensemble forecasts for weather and seasonal climate, *J. Clim.*, *13*, 4196–4216.
- Kug, J. S., I. S. Kang, and D. H. Choi (2008a), Seasonal climate predictability with tier-one and tier-two prediction system, *Clim. Dyn.*, *31*, 403–416.
- Kug, J. S., J. Y. Lee, I. S. Kang, B. Wang, and C. K. Park (2008b), Optimal multi-model ensemble method in seasonal climate prediction, *J. Kor. Meteor. Soc.*, *44*, 259–267.
- Kug, J. S., J. Y. Lee, and I. S. Kang (2008c), Systematic error correction of dynamical seasonal prediction of sea surface temperature using a stepwise pattern project method, *Mon. Weather Rev.*, *136*, 3501–3512.
- Kurihara, Y., M. A. Bender, R. E. Tuleya, and R. J. Ross (1995), Improvements in the GFDL hurricane prediction system, *Mon. Weather Rev.*, *123*(9), 2791–2801.
- Lee, Y. H., S. K. Park, and D. E. Chang (2006), Parameter estimation using the genetic algorithm and its impact on quantitative precipitation forecast, *Ann. Geophys.*, *24*, 3185–3189.
- Lu, C., H. Yuan, B. E. Schwartz, and S. G. Benjamin (2007), Short-range numerical weather prediction using time-lagged ensembles, *Weather Forecasting*, *22*(3), 580–595.
- Michaelsen, J. (1987), Cross-validation in statistical climate forecast models, *J. Climate Appl. Meteorol.*, *26*(11), 1589–1600.
- Min, Y. M., V. N. Kryjov, and S. M. Oh (2014), Assessment of APCC multimodel ensemble prediction in seasonal climate forecasting: Retrospective (1983–2003) and real-time forecasts (2008–2013), *J. Geophys. Res. Atmos.*, *119*, 12,132–12,150, doi:10.1002/2014JD022230.
- Nasseri, M., K. Asghari, and M. J. Abedini (2008), Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network, *Expert Syst. Appl.*, *35*, 1415–1421.
- Oh, S.-G., and M.-S. Suh (2016), Comparison of projection skills of deterministic ensemble methods using pseudosimulation data generated from multivariate Gaussian distribution, *Theor. Appl. Climatol.*, 1–20, doi:10.1007/s00704-016-1782-1.
- Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délecluse, and M. Déqué (2004), Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER), *Bull. Am. Meteorol. Soc.*, *85*(6), 853–872.
- Park, S. H., S. D. Kang, and W. T. Kwon (2005), Prediction of boreal winter precipitation by nonlinear multimodel ensemble technique, *J. Kor. Meteor. Soc.*, *41*, 1015–1028.
- Peng, P., A. Kumar, H. Van den Dool, and A. G. Barnston (2002), An analysis of multimodel ensemble predictions for seasonal climate anomalies, *J. Geophys. Res.*, *107*(D23), 4710, doi:10.1029/2002JD002712.
- Saha, S., S. Nadiga, C. Thiaw, J. Wang, and W. Wang (2006), The NCEP climate forecast system, *J. Clim.*, *19*, 3483–3517.
- Stensrud, D. J., H. E. Brooks, J. Dun, M. S. Tracton, and E. Rogers (1999), Using ensembles for short-range forecasting, *Mon. Weather Rev.*, *127*(4), 433–446.
- Stensrud, D. J., J. W. Bao, and T. T. Warner (2000), Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems, *Mon. Weather Rev.*, *128*(7), 2077–2107.
- Suh, M. S., S. G. Oh, D. K. Lee, D. H. Cha, S. J. Choi, C. S. Jin, and S. Y. Hong (2012), Development of new ensemble methods based on the performance skills of regional climate models over South Korea, *J. Clim.*, *25*(20), 7067–7082.
- Sun, J., and J. B. Ahn (2011), A GCM-based forecasting model for the landfall of tropical cyclones in China, *Adv. Atmos. Sci.*, *28*, 1049–1055.
- Sun, J., and J. B. Ahn (2014), Dynamical seasonal predictability of the arctic oscillation using a CGCM, *Int. J. Climatol.*, *35*(7), 1342–1353.
- Tabassum, M., and K. Mathew (2014), A genetic algorithm analysis towards optimization solutions, *Int. J. Digit. Inf. Wirel. Commun.*, *4*(1), 124–142.
- Van den Dool, H. M., and L. Rukhovets (1994), On the weights for an ensemble-averaged 6–10 day forecast, *Weather Forecasting*, *9*(3), 457–465.
- Wang, B., et al. (2009), Advance and prospectus of seasonal prediction: Assessment of the APCC/ClipAS 14-model ensemble retrospective seasonal prediction (1980–2004), *Clim. Dyn.*, *33*, 93–117, doi:10.1007/s00382-008-0460-0.
- Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences*, 467 pp., Academic Press, New York.
- You, S. H., Y. H. Lee, and W. J. Lee (2012), Parameter estimations of a storm surge model using a genetic algorithm, *Nat. Hazards*, *60*(3), 1157–1165.
- Yun, W. T., and T. N. Krishnamurti (2002), Linear and non-linear multi-model superensemble prediction model, *J. Kor. Meteor. Soc.*, *12*, 26–31.
- Yun, W. T., L. Stefanova, and T. N. Krishnamurti (2003), Improvement of the superensemble technique for seasonal forecasts, *J. Clim.*, *16*, 3834–3840.
- Yun, W. T., W. S. Lee, and T. N. Krishnamurti (2005), Seasonal prediction of precipitation using multi-model synthetic superensemble algorithm, *J. Kor. Meteor. Soc.*, *41*, 159–172.
- Yuval, and W. W. Hsieh (2003), An adaptive nonlinear MOS scheme for precipitation forecasts using neural networks, *Weather Forecasting*, *18*(2), 303–310.
- Zhong, A., H. H. Hendon, and O. Alves (2005), Indian Ocean variability and its association with ENSO in a global coupled model, *J. Clim.*, *18*, 3634–3649.