INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **35**: 3721–3732 (2015)
Published online 5 December 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4232

RMetS
Royal Meteorological Society

Improvement of CGCM prediction for wet season precipitation over Maritime Continent using a bias correction method

Sera Jo and Joong-Bae Ahn*

Division of Earth Environmental System, Pusan National University, South Korea

ABSTRACT: A new model output statistics method – Ensemble Selective Simple Linear Regression (E-SSLR) – is developed based on SLR in order to increase the seasonal prediction skill of a Coupled General Circulation Model (CGCM) over the Maritime Continent (MC), a region with large model simulation errors. E-SSLR is applied to Pusan National University (PNU) CGCM hindcast over the MC region for the period of 1981–2010 to reduce the systematic model bias in boreal winter (DJF) seasonal mean precipitation and outgoing long-wave radiation (OLR) anomalies. Three oceanic indices (Nino 3.4, El Nino Modoki and Indian Ocean Dipole (IOD) Mode indices) and one atmospheric index (Southern Oscillation Index, SOI) produced from PNU CGCM hindcast are used as SLR predictor. E-SSLR consists of three steps: Selection, SLR and Ensemble. The selection and ensemble steps are added to the conventional SLR step to overcome the weakness of the linear regression method. In the selection step, the grids with a temporal correlation coefficient between predictor and predictand exceeding the threshold are selected. These grids (grid-selected) are corrected by SLR in the second step. For the grids that are grid-not-selected, the original CGCM results are used without further correction. This prevents insignificant statistical correction due to the application of low correlated predictors to the SLR. The correction effect of E-SSLR is analysed in terms of deterministic and categorical analyses. The result shows that the seasonal predictability of DJF seasonal precipitation and OLR in the MC region is increased by using E-SSLR, and this increment is statistically significant. The correction effect is larger when indices with high predictability that are closely correlated with the predictand are used as predictors.

KEY WORDS Maritime Continent precipitation; model bias correction; CGCM predictability; model output statistics; simple linear regression

Received 6 May 2014; Revised 29 September 2014; Accepted 10 November 2014

1. Introduction

The Maritime Continent (MC), one of the world's regions of active convection and heavy precipitation, lies between the tropical western Pacific and the eastern Indian Ocean, so-called warm pool area. Extreme events such as droughts, floods and typhoons are very frequent over south-east Asian countries such as Indonesia, Malaysia, Vietnam, Cambodia, Philippines and Papua New Guinea. For example, during 2007 winter, Indonesia experienced heavy flooding and recorded 112 casualties because of continuous heavy rainfall. In 2011 fall, flooding due to precipitation lasted almost 4 months caused over 1000 casualties and huge economic damage throughout south-east Asia, including Thailand, Cambodia, Vietnam and Philippines. On the contrary, severe drought led to crop failure in 2005 and 2010, and huge forest fires in Borneo Island and some parts of Indonesia and Malaysia in 1997 and 1998. During these periods, enormous loss of forest and severe economic damage were suffered. Thus, accurate long-term prediction in this MC is crucial to cope with natural disasters.

This warm pool region plays an important role in global circulation as rising branches of Hadley and Walker circulations and Indian Ocean Dipole (IOD). The warmth of the sea surface temperature (SST) in this area leads to a huge amount of surface heat flux associated with active evaporation resulting in vigorous deep convection (Ramage, 1968). The precipitation induced by such active convection leads to a huge amount of latent heat release at the middle of troposphere, which drives a rising motion that affects the global-scale climate through Hadley and Walker circulations and IOD. Ramage (1968) named this region the 'boiler box' of the earth. Besides, the recent study of Sun *et al.* (2009) found that the convection activity over the region of the MC serves as a bridge linking the boreal spring Antarctic Oscillation and the Yangtze River valley summer rainfall. This implies that MC plays an important role in interaction between the Southern–Northern Hemispheres. In spite of its climatological importance, the Coupled General Circulation Model (CGCM) has huge bias over this region in precipitation (Neale and Slingo, 2003;

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

Qian, 2008; Schiemann *et al.*, 2014). Because of complex orography and topography with over 3000 islands of various size surrounded by shallow and narrow seas, prediction and simulation of weather and climate over this area with various scales of space and time are challenging works.

Neale and Slingo (2003) insisted that systematic error of General Circulation Model (GCM) in the MC eventually affects the simulation of general circulation as convection in this region strongly influences global-scale energy and moisture budgets, thereby decreasing the GCM's predictability in higher latitudes. Therefore, it is important to increase the model predictability over the MC area not only for this region but also for higher latitudes. The low predictability over this region results from model error or bias, which can be divided into two parts: systematic bias and non-systematic errors. The systematic bias can be further subdivided into the mean and anomalous parts (Kug *et al.*, 2007; Jin *et al.*, 2008; Wang *et al.*, 2008; Ahn *et al.*, 2012). The systematic mean bias can be corrected relatively easily by removing the difference in climatology between model and observation. However, the anomalous bias is difficult to correct because of its irregular pattern or trend.

Model output statistics (MOS; Glahn and Lowry, 1972; Wilks, 1995) is a kind of statistical post-processing methods to improve the skill of model prediction. One of the most common MOS methods is a simple linear regression (SLR) method. This method uses a linear relationship between the target variable and other variables, including meteorological indices and patterns such as El Nino/Southern Oscillation (ENSO), Arctic Oscillation and Pacific-North America. Different kinds of MOS methods use the pattern relationship between model and observation. These methods usually utilize singular value decomposition and empirical orthogonal function analyses in finding the patterns. Non-linear methods such as artificial neuron network and genetic algorithm are also used for MOS.

In this study, SLR is used to reduce the anomalous bias of seasonal hindcast results from a CGCM and increase the long-term predictability of precipitation and outgoing long-wave radiation (OLR) over the MC region in the wet season. The precipitation and OLR represent the local-scale and large-scale convective activities, respectively. As for the predictors of the regression equation, the large-scale tropical indices strongly affecting precipitation and convection in the MC region are considered. Especially, Nino 3.4, El Nino Modoki and IOD Mode indices, represented by equatorial SST, which has relatively good predictability in most CGCMs, are chosen as predictors to correct the precipitation bias over the region.

2. Data and Method

2.1. Data

This study uses hindcast results of the Pusan National University (PNU) CGCM for the application of a statistical correction method to rainy season precipitation over the MC area. The PNU CGCM is a participant model of Asia-Pacific Economic Cooperation Climate Center (APCC) multi-model ensemble long-range prediction system (Sun and Ahn, 2011, 2014). This model describes interactions between four subsystems: atmosphere, land-surface, ocean and sea-ice. The atmospheric component model of this CGCM is the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM3; Hurrell *et al.*, 1998) with a triangular truncation at wavenumber 42 and 18 vertical levels. The Land Surface Model is combined with atmospheric GCM. The oceanic component model is Geophysical Fluid Dynamics Laboratory Modular Ocean Model version 3 (MOM3; Pacanowski and Griffies, 1999). For the sea ice, the elastic viscous plastic sea-ice model (Los Alamos National Laboratory) modified by Ahn and Lee (2001) is coupled with the atmospheric and oceanic models. No flux adjustment is applied in the experiment. The hindcast consisted of five ensemble members using time-lag method. In this study, a simply composited ensemble mean is used for the statistical correction. The PNU CGCM prescribes data-assimilated ocean initial condition using the variational method using filter (Huang, 2000). More detailed explanation is presented by Sun and Ahn (2011, 2014) and Ahn and Kim (2014). The CGCM hindcasts with October initial condition 0–4 month lead time and seasonal mean [lead2 (December)–lead4 (February), hereafter DJF] are used to make predictors to correct for the MC wet season precipitation and OLR. Boreal winter (DJF) is regarded as the wet season of the MC region. In order to set an SLR equation and validate the results, Global Precipitation Climatology Project v2.2 (GPCP; Adler *et al.*, 2003) precipitation and National Centers for Environmental prediction (NCEP) and NCAR Reanalysis 2 (NCEP/NCAR R2; Kanamitsu *et al.*, 2002) OLR data are used. The horizontal resolutions of the data are $2.5° \times 2.5°$ in both longitudinal and latitudinal directions.

2.2. Method

This study applies an Ensemble Selective Simple Linear Regression (E-SSLR) method to the CGCM hindcast for the 30-year period of 1980–2009. Since the sample period is not long enough to be sub-divided into training and validating periods, leave-one-out cross-validation (Michaelsen, 1987; Barnston, 1994) is used by taking 29 years for the training period and 1 year for forecast. At each forecast process (30 times), we apply three steps – selection, SLR and ensemble – independently. Figure 1 shows a flow chart of the SSLR method. First, we set a threshold which is defined as a critical value at the 90% confidence level in the temporal correlation coefficient (TCC) between indices (predictor) and target variable (predictand). After setting the threshold, we select the grids with a TCC exceeding the threshold as 'grid-selected'. After selection step, only 'grid-selected' are corrected by SLR in second step. For the grids that are 'grid-not-selected', the original CGCM results are used without further correction. The intention of selection step before SLR is to retain the predictability of

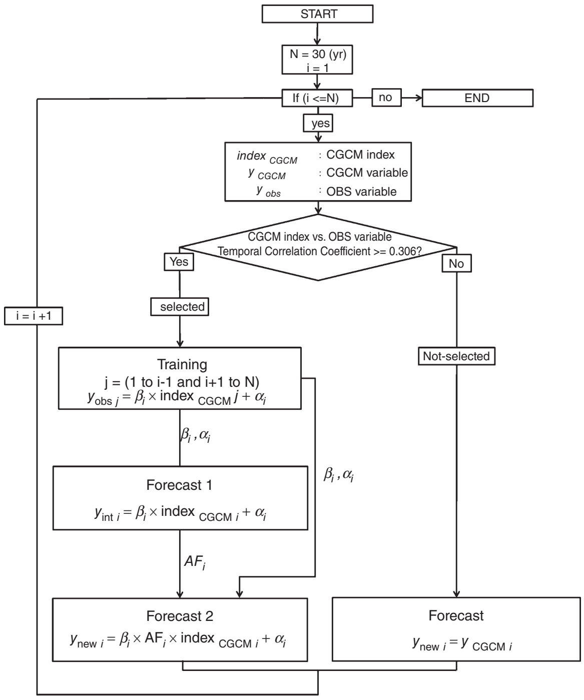

Figure 1. Flow chart of SSLR.

the CGCM results, which are based on dynamics, when the relationship between predictor and the predictand is not prominent. The SLR correction equation for the training and validation at the selected grids for each cross-validation step i is given as follows.

$$\text{Training}: \; y_{\text{obs}\,j} = \beta_i \times \text{index}_{\text{CGCM}\,j} + \alpha_i \quad (i,j = 1 \sim N,\; j \neq i) \tag{1}$$

$$\text{Forecast 1}: \; y_{\text{int}\,i} = \beta_i \times \text{index}_{\text{CGCM}\,i} + \alpha_i \quad (i = 1 \sim N) \tag{2}$$

$$\text{Forecast 2}: \; y_{\text{new}\,i} = \beta_i \times AF_i \times \text{index}_{\text{CGCM}\,i} + \alpha_i \quad (i = 1 \sim N) \tag{3}$$

$$AF_i = {}^{\sigma_{\text{obs}\,i}}/_{\sigma_{\text{int}\,i}} \quad (i = 1 \sim N) \tag{4}$$

$N = 30$ (years), and the β_i and α_i are regression coefficients that satisfy the least-square method. The standard deviations, $\sigma_{\text{int}\,i}$ and $\sigma_{\text{obs}\,i}$, are defined as

$$\sigma_{\text{int}\,i} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(y_{\text{int}\,j} - \overline{y_{int\,i}}\right)^2} \quad (i,\, j = 1 \sim N,\, j \neq i) \tag{5}$$

$$\sigma_{\text{obs}\,i} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(y_{\text{obs}\,j} - \overline{y_{\text{obs}\,i}}\right)^2} (i,\, j = 1 \sim N,\, j \neq i) \tag{6}$$

In addition, the average of x_i is calculated as:

$$\overline{x_i} = \frac{1}{N-1}\sum_{j=1}^{j=N} x_j \quad (i,\, j = 1 \sim N,\, j \neq i) \tag{7}$$

In the training equation (1), observation data are used for the predictand variable. The independent variable is index that is predicted by CGCM. Using the linear relationship between the dependent and independent variables, the regression coefficients β_i and α_i are determined for each target year, i. This means regression coefficients are different for each validation year. After 'training step', we make intermediate SLR result ($y_{\text{int}\,i}$) for 30 years of precipitation and OLR using Equation (2). The standard deviation of $y_{\text{int}\,i}$ ($\sigma_{\text{int}\,i}$) is calculated from 30 years of $y_{\text{int}\,i}$ with β_i and α_i of each target year i (Equation (5)). Also, 30 different standard deviation of observation ($\sigma_{\text{obs}\,i}$) are calculated using leave-one-out data (6). Finally, the corrected value for ith year ($y_{\text{new}\,i}$) is determined in Equation (3). Here, β_i is multiplied by AF$_i$ which is 'amplification factor'. This AF$_i$ is the ratio of $\sigma_{\text{obs}\,i}$ to $\sigma_{\text{int}\,i}$ (Equation (4)). The role of AF$_i$ is solving the problem in that the standard deviation is decreased after statistical correction relative to the observation which generally incurs. After SSLR with 18 predictor indices (Nino 3.4, EM and IODM indices with different lead times), final correction result (E-SSLR) is obtained using simple composite method (SCM) in step 3.

2.3. Experiment design

The prediction of MC precipitation is corrected using E-SSLR, which utilizes CGCM-predicted indices as predictor. As the equatorial SST is relatively well simulated compared with other variables, it has often been used as a predictor for MOS methods or statistical models in many studies (Barnston *et al.*, 1996; Kug *et al.*, 2008). PNU CGCM also has good predictability in terms of equatorial Pacific Ocean SST particularly in Nino 3.4 region (Jeong and Ahn, 2007). Therefore, SST-based indices produced by the CGCM are chosen as predictors of SLRs to correct for hindcasted precipitation and OLR predictions. In this study, the MC is defined as the area of 90–160°E longitudinally and 15°S–15°N latitudinally (Figure 2(a)). The indices used in this study are Nino 3.4 index (Nino3.4I), Southern Oscillation Index (SOI), El Nino Modoki Index (EMI) and IOD Mode Index (IODMI), as shown in Table 1. The target period for the prediction and correction is boreal winter (DJF seasonal mean), which is the wet season in the area (Figure 2(b)).

The convection and rainfall activity over the MC has relatively good relationship with predictor indices during the wet season. That is, in the year of El Nino, the MC area has colder SST than normal, resulting in suppressed convection and rainfall activity (Lau and Chan, 1983; McBride *et al.*, 2003; Chang *et al.*, 2004). Therefore, Nino 3.4I and SOI, which represent the ENSO phenomena, are reasonable predictors for MC precipitation. EMI, which represents a different kind of ENSO phase, is also selected as a predictor. El Nino Modoki is a phenomenon characterized by warm SST anomalies over the Central

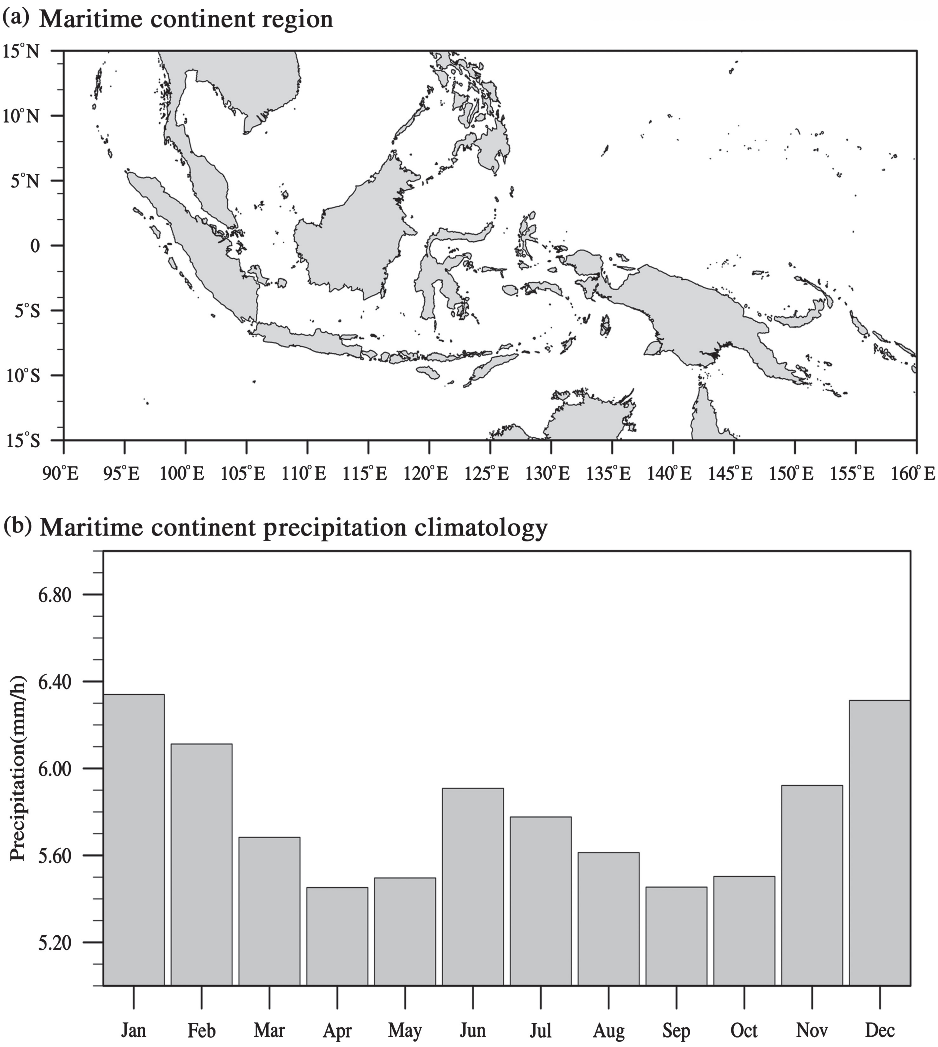

Figure 2. (a) MC region and (b) precipitation climatology over the area.

Pacific and cold over the western and eastern Pacific. Another air–sea interaction that influences the target region is IOD, introduced by Behera *et al.* (1999) and Webster *et al.* (1999). IOD is a phenomenon with a seesaw pattern of SST anomalies between the eastern and western Indian Ocean. The negative phase of IOD brings warmer SST condition around the eastern Indian Ocean, inducing increased precipitation over MC, including Indonesia. Thus, IODMI (Saji *et al.*, 1999) is also taken into account as a predictor for precipitation correction. SOI, the atmospheric index, is selected as a predictor for comparison with the oceanic index, Nino 3.4I.

Figures 3 and 4 show the TCCs between DJF seasonal mean precipitation and OLR, and four kinds of indices from October to February. The regions with confidence level over 90% are shaded. Nino 3.4I and SOI for DJF show high correlation coefficients with MC precipitation and OLR. All the indices of DJF show negative correlation with precipitation activity. This implies that the negative phases of ENSO, IOD and El Nino Modoki lead warmer SST condition around the MC region, inducing more vigorous convection and precipitation. During October to

Table 1. Indices used for the correction of MC precipitation and OLR.

Index	Description
Nino 3.4I	Area-averaged SSTA in 5°–5°N, 190°–240°E
EMI	Area-averaged SSTA in A =165°–220°E, 10°–10°N B =250 –290°E, 15°–5°N C =125 –145°E, 10°–20°N. EMI = A – (0.5*B) – (0.5*c)
IODMI	Area-averaged SSTA in A = 50 –70°E, 10°–10°N B = 90 –110°E, 10°–10°N. IODMI = A–B
SOI	mslp(Thaiti) – mslp(Darwin)

February, Nino 3.4I is negatively correlated with DJF precipitation. SOI shows the same result with Nino 3.4I but with the opposite sign. These results are similar to those by Chang *et al.* (2004). During October to February, EMI also shows a negative relationship with DJF precipitation over the limited area of the Philippines and center of the MC Ocean. On the other hand, IODMI shows a 3- to

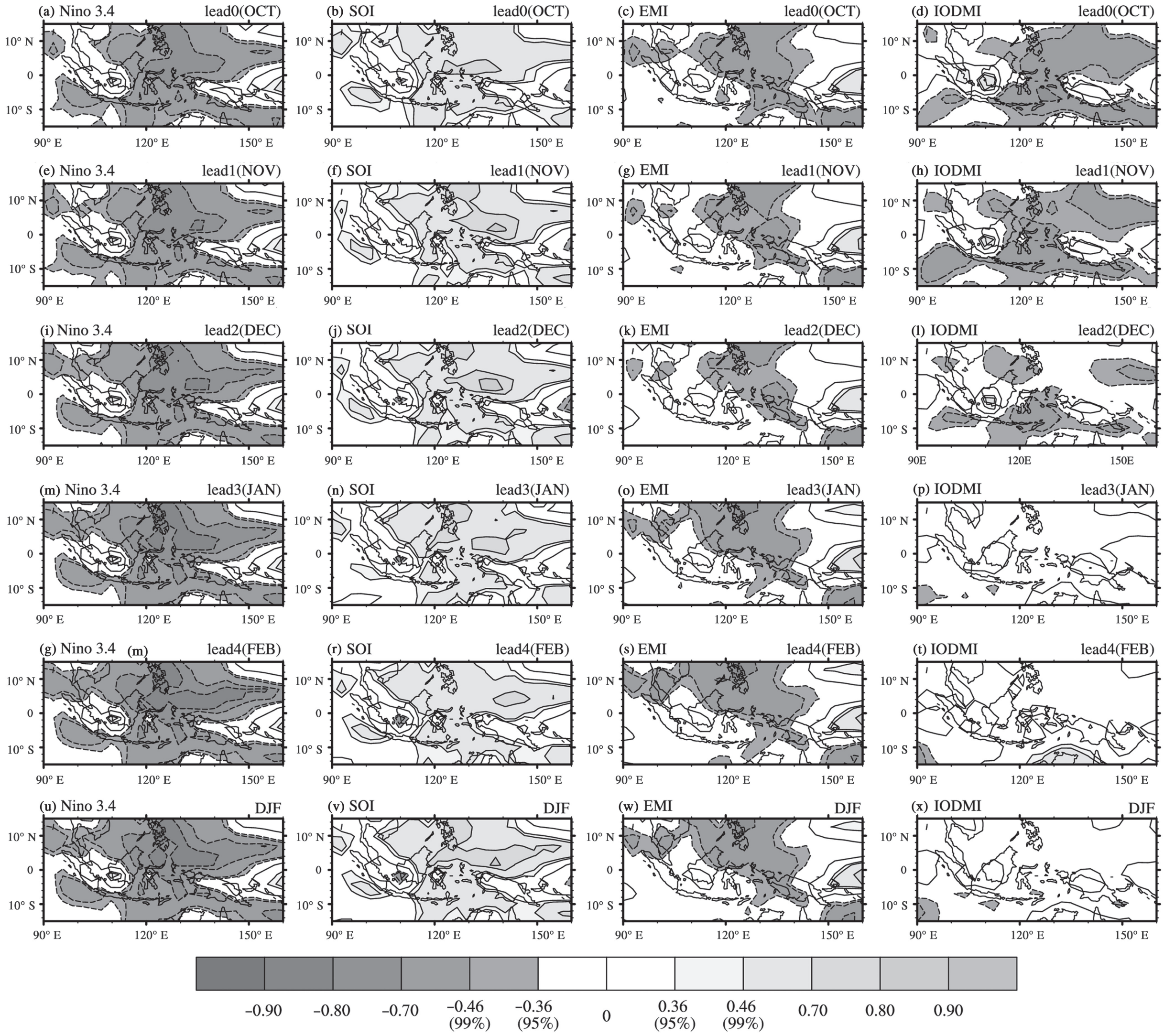

Figure 3. TCC between observed precipitation and predictor indices.

5-month lagged correlation with a significant correlation only in October–December. Figures 3 and 4 show that the predictor indices are appropriate to correct precipitation and OLR over the MC region.

The TCCs between the indices from the CGCM hindcast and observation are shown in Table 2 to illustrate the ability of the CGCM to predict the predictor indices. Nino 3.4I maintains the predictability from leads 0 to 4 with a TCC over 0.9. This indicates that PNU CGCM can replicate the ENSO phenomenon closely. SOI and IODMI show a decline in predictability with increasing lead time. Nonetheless, predictability of all the monthly indices as well as that of seasonal mean indices remains significant with exception of lead4 of SOI and IODMI, which means that those four indices can profitably be used as predictor for SLR. Furthermore, Table 2 reveals that the SST-based indices (Nino 3.4I, EMI and IODMI) are more accurately produced than the atmospheric variable (mean sea-level pressure)-based index (SOI), which suggests that SST-based indices will improve the correction result. Each of the 24 predictors from the four indices of the six lead times is applied independently to the SSLR method introduced in Section 2.2. To generate the E-SSLR result, we apply SCM to the SSLR of the oceanic indices (Nino3.4I, EMI and IODMI).

Section 3 presents the correction results according to deterministic analysis and categorical deterministic analysis.

3. Results

3.1. Deterministic predictions of anomalies

The prediction skill of E-SSLR result is analysed by deterministic analysis to investigate the improvement in the DJF seasonal prediction of precipitation and OLR. Figure 5 shows TCC and root mean square error (RMSE) of the hindcasted (Figure 5(a) and (c)) and corrected (Figure 5(b) and (d)) precipitation. The shaded areas in Figure 5(a) and (b) show statistically significant TCCs at the 95% confidence level. The dotted area in Figure 5(b) represents

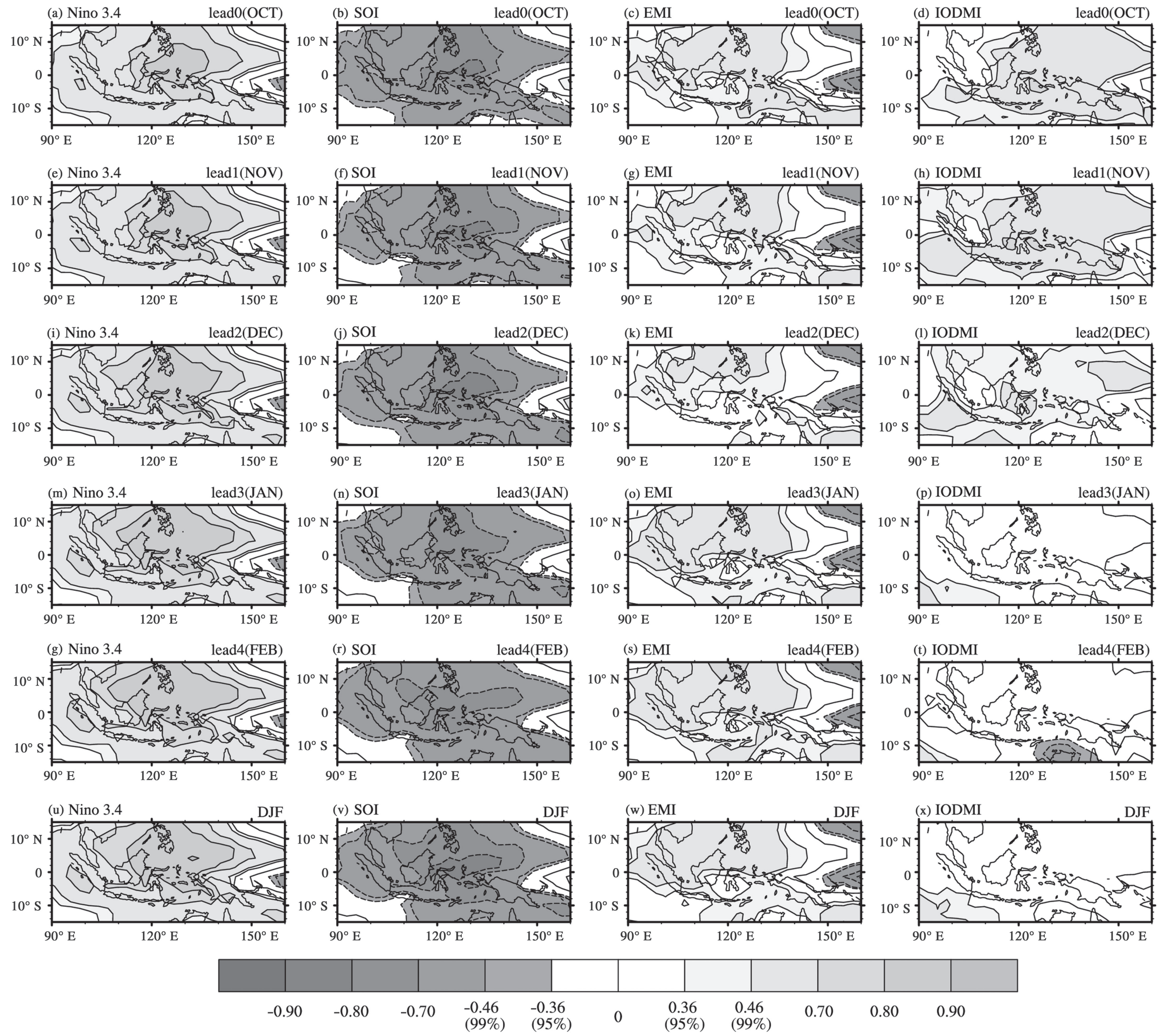

Figure 4. TCC between observed OLR and predictor indices.

Table 2. The TCC between the indices of the CGCM hindcast with different lead times and observation.

	Lead0 (October)	Lead1 (November)	Lead2 (December)	Lead3 (January)	Lead4 (February)	DJF
Nino 3.4I	0.96**	0.95**	0.93**	0.93**	0.89**	0.94**
EMI	0.93**	0.83**	0.70**	0.77**	0.73**	0.75**
IODMI	0.92**	0.88**	0.71**	0.47**	0.10	0.52**
SOI	0.82**	0.75**	0.79**	0.56**	0.09	0.63**

**: 99% confidence level.

grids with TCCs that are not significant at the 99% confidence level in hindcast, but that become significant at the 99% confidence level after correction. The uncorrected precipitation and OLR have TCCs that are not significant at the 95% confidence level over the whole domain, except the southern part of the Philippines and the narrow eastern part of Papua New Guinea. RMSE is relatively higher in areas of lower TCC than in other regions, especially in Papua New Guinea and Borneo Island (Figure 5(c)). However, after correction, a broad area of the MC region shows significant TCC at the 95% confidence level and the area-averaged value also shows significant TCC at the 99% confidence level (Figure 5(b)). The area-averaged RMSE also decreases from 0.89 to 0.75, and the TCC increment is large over the western Pacific (Figure 5(d)). The improvement is clear in the case of OLR, as shown in Figure 6. The OLR shows statistically significant prediction skill at the 95% confidence level over a broad area

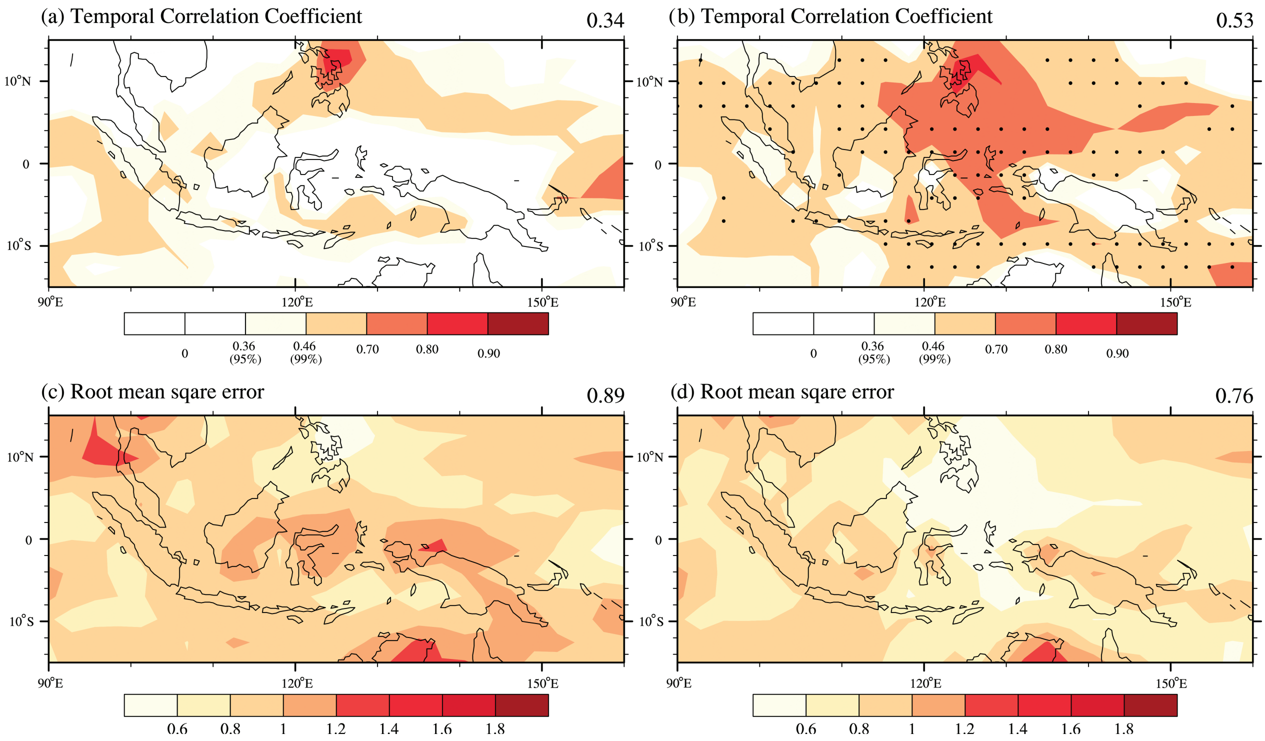

Figure 5. (a) and (b) are the TCC of hindcasted and corrected precipitation. The dotted area represents the grids where the significance of the TCC increased above the 99% confidence level after correction. (c) and (d) are the normalized RMSE of hindcasted and corrected precipitation.

in Figure 6(a) as the CGCM predicts OLR better than precipitation. However, it still shows a low TCC, which is not significant at the 95% confidence level around Papua New Guinea. Nevertheless, as shown in Figure 6(b), the OLR of the overall MC region shows significant TCCs at the 95% confidence level after correction. According to the results of Figures 5 and 6, the correction effect is clear in precipitation. Of total 250, 105 grids become significant at 99% level of confidence after correction. OLR is also corrected significantly at 82 grids (Figure 6(b)). As TCC of OLR is higher than precipitation before correction, it has relatively less chance to be improved than precipitation. That is, while the number of grid improved is larger in precipitation, number of grid with TCC over 99% confidence level is larger in OLR. The area-averaged TCC also increases from 0.45 to 0.53. Figure 6(c) and (d) shows the RMSE decrease, especially at the central part of the MC region. The area-averaged RMSE also decreases from 0.81 to 0.77, which shows that the MC region hindcast error is decreased by the correction.

Figure 7(a) and (c) presents the Taylor diagram (Taylor, 2001) and pattern correlation coefficient (PCC)–RMSE scatterplot for precipitation, respectively, and Figure 7(b) and (d) does the same for OLR. The dotted lines in the Taylor diagrams represent TCCs significant at the 95% and 99% confidence level, and each dot denotes the area-averaged value of the MC domain. The black dot represents the predictability of hindcasted precipitation, and each of SSLR results with CGCM Nino 3.4I, EMI, IODMI and SOI and are represented by the different shapes of dots (open circle, triangle, square and asterisk, respectively). The ensemble members of E-SSLR are SSLR results with Nino3.4I, EMI and IODMI, except SOI. However, we present SOI SSLR results (marked with asterisk) to compare oceanic and atmospheric indices. The E-SSLR result is marked with red star. As shown in Figure 7(a), the hindcasted precipitation is not statistically significant at the 95% level, but the SSLR results using Nino 3.4I, EMI and IODMI as predictor have TCC over the 95% confidence level. However, the correction with SOI as predictor shows lower performance than that with Nino 3.4I, although El Nino and Southern Oscillation are merely different aspects of the same phenomenon. This is because SOI is less predictable than Nino 3.4I by CGCM (Table 2). The red star that represents the E-SSLR shows the highest TCC, over 99% confidence level. The PCC–RMSE scatterplot of precipitation shows the trend of increasing PCCs and decreasing RMSEs for each ensemble members of E-SSLR and E-SSLR, except for SSLR with SOI (Figure 7(c)). This suggests that the correction improves the predictability of spatial distribution of precipitation. Figure 7(b) and (d) is the same as Figure 7(a) and (c), but for the OLR rather than precipitation. The confidence level of TCC for OLR is improved from 95 to 99%. In addition, PCC increases and RMSE decreases more in OLR than in the case of precipitation.

3.2. Deterministic predictions of categories

In this section, the seasonal predictability with regard to hindcasted and corrected precipitation (and OLR) is evaluated using the categorical deterministic forecast. The categorical deterministic forecast classifies the forecast into

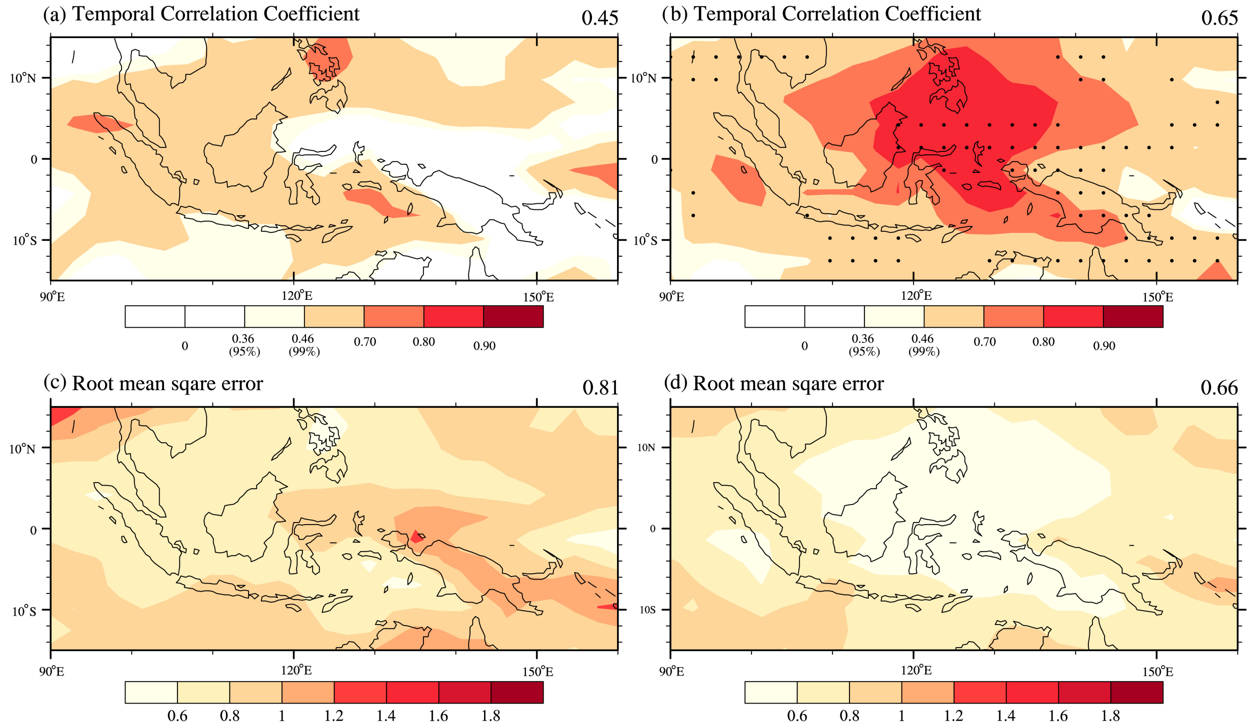

Figure 6. (a) and (b) are the TCC of hindcasted and corrected OLR. The dotted area represents the grids where the significance of the TCC increased above the 99% confidence level after correction. (c) and (d) are the normalized RMSE of hindcasted and corrected OLR.

three categories with respect to climatology: above normal, normal and below normal. In this study, the threshold of each category is set to times the standard deviation that has a population ratio of 3:4:3 with respect to above normal, normal and below normal categories, assuming that the probability distribution of precipitation (or OLR) follows a Gaussian distribution. The categorical deterministic forecast is generally used for long-term and seasonal predictions, for which precise prediction of anomaly values is difficult. In this study, the predictability of the categorical deterministic forecast is evaluated in terms of hit rate (HR), false alarm rate (FAR) and Heidke Skill Score (HSS) using a 3×3 contingency table (Table 3). The three columns of Table 3 are category for forecast, and rows are for observation. The '+', '−' and '0' denote anomalies exceeding +0.53 (above normal), below −0.53 (below normal) and between −0.53 and 0.53 (normal) of standard deviation. Therefore, each event can be classified into nine categories. As an example, if model forecast of DJF precipitation of a year is above normal and observed precipitation is also above normal, the year is categorized into category A, etc. HR, FAR and HSS are calculated using the following equations (Wilks, 1995).

$$\mathrm{HR} = \frac{\mathrm{A+F+K}}{\mathrm{P}} \tag{8}$$

$$\mathrm{FAR} = \frac{(\mathrm{E+I})+(\mathrm{B+J})+(\mathrm{C+G})}{(\mathrm{P-D})+(\mathrm{P-H})+(\mathrm{P-L})} \tag{9}$$

$$\mathrm{HSS} = \frac{(\mathrm{A+F+K})-\mathrm{C_1}}{\mathrm{P-C_1}} \tag{10}$$

$$\mathrm{C_1} = 0.3 \times (\mathrm{M+O}) + 0.4 \times N \tag{11}$$

Figure 8(a), (c) and (e) shows HR, HSS and FAR of DJF seasonal mean hindcasted precipitation of PNU CGCM, while Figure 8(b), (d) and (f) shows HR, HSS and FAR after correction, respectively. With regard to the MC area-averaged value, HR and HSS increase from 0.45 to 0.51 and from 0.16 to 0.28, while FAR decreases from 0.27 to 0.24, indicating that the correction result of precipitation using E-SSLR increases the predictability of the categorical deterministic forecasts. Figure 9, which is the same as Figure 8, shows HR, HSS and FAR with regard to seasonal mean OLR before and after correction. While OLR has higher HR and HSS than those of precipitation before correction, it is further improved after correction so that area-averaged HR and HSS increase from 0.48 to 0.59 and from 0.22 to 0.38, while FAR decreases from 0.26 to 0.20. Figure 10 shows a scatterplot with the x-axis of FAR and the y-axis of HR, in which each point represents an area-averaged value over the MC region. The legend in Figure 10 is the same as that in Figure 7. Figure 10(a) is a scatterplot of precipitation, which shows that the categorical deterministic forecast ability is improved as the HR of all ensemble members and E-SSLR increases while FAR decreases compared with the black point. In the case of OLR, the correction is also well applied, although the HR increment and FAR decrement are lower than that of the precipitation (Figure 10(b)). The correction effect of E-SSLR is analysed via deterministic analysis and categorical deterministic analysis, and the result shows that the seasonal predictability is increased by E-SSLR with regard to DJF seasonal precipitation and OLR in the

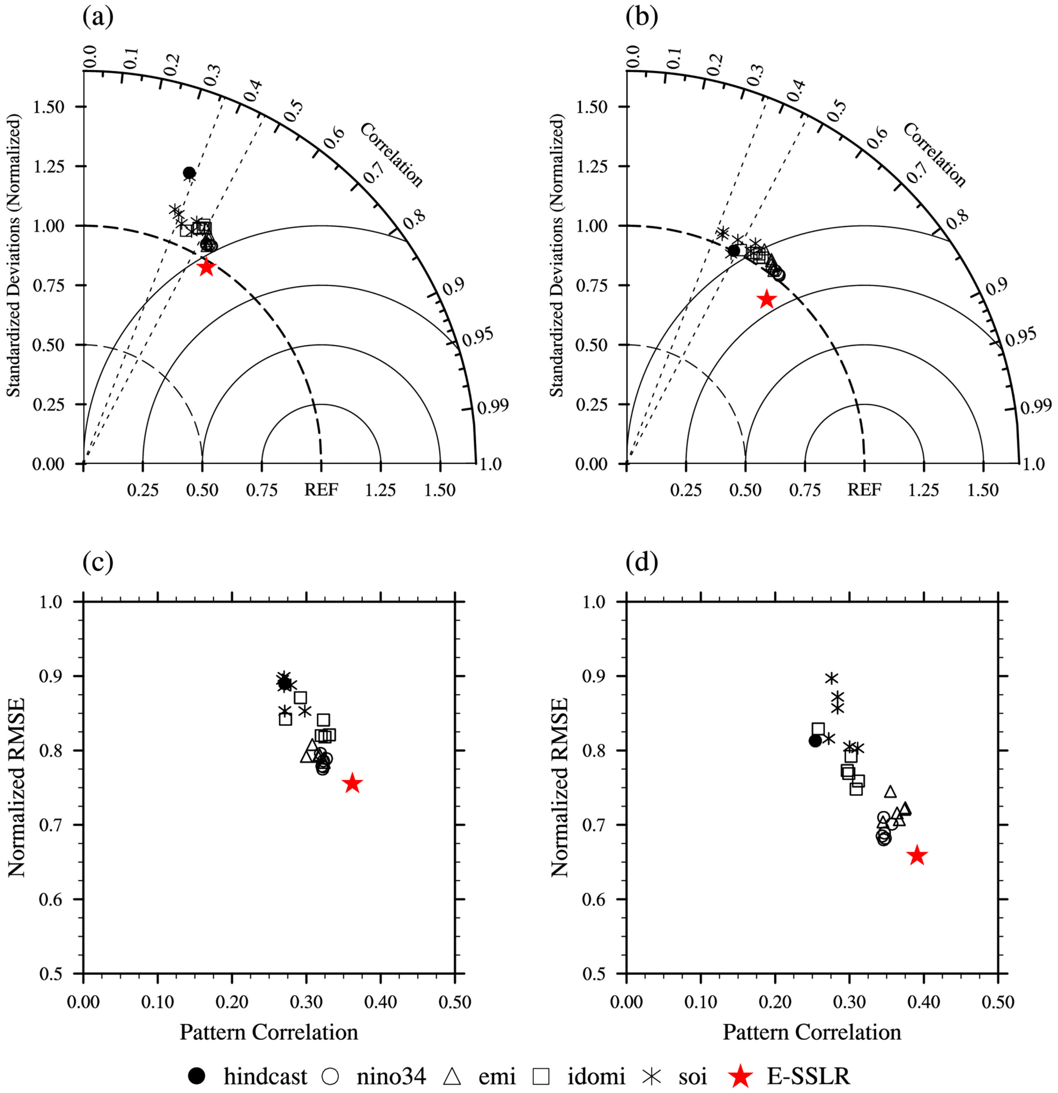

Figure 7. Taylor diagram of (a) precipitation and (b) OLR, and PCC-RMSE scatterplot of (c) precipitation and (d) OLR.

Table 3. Contingency table (3 × 3) for categorical forecast.

FCST and OBS	+	0	−	Total
+	A	B	C	D
0	E	F	G	H
−	I	J	K	L
Total	M	N	O	P

FCST and OBS indicate the forecast and observation categorized, respectively, as above normal (+), normal (0) and below normal (−).

MC region. Furthermore, the effect of correction is larger when indices with high predictability closely correlated with predictand are used as predictors.

4. Summary and conclusion

Many systematic bias occur in the MC region of the CGCM, especially for precipitation. This study proposes a statistical correction method to increase the predictability of DJF seasonal precipitation and OLR in the MC region simulated by PNU CGCM. The precipitation in the MC region is corrected using indices derived from the indices of ENSO, El Nino Modoki and IOD phenomena, which are simulated relatively well by the dynamical model and also show high correlation with precipitation in the MC region. The indices are produced from the hindcast SST of the model to enable application of this method to seasonal forecasts. In addition, SOI, which is produced using the hindcasted MSLP, is also used as a predictor to determine the effect of the accuracy of the model indices on the correction. The designed correction method is E-SSLR, which is a modification of SLR, the simplest and most widely used linear regression method. E-SSLR consists of three steps: selection, SLR and ensemble. The SLR step is the same as the commonly used SLR, while the other two steps are added to overcome the weakness of the linear regression method. Through the selection step, SLR is selectively applied only for selected grids which have high correlation coefficients between precipitation (or OLR) and predictor index. This procedure prevents insignificant statistical correction due to application of low correlated predictors

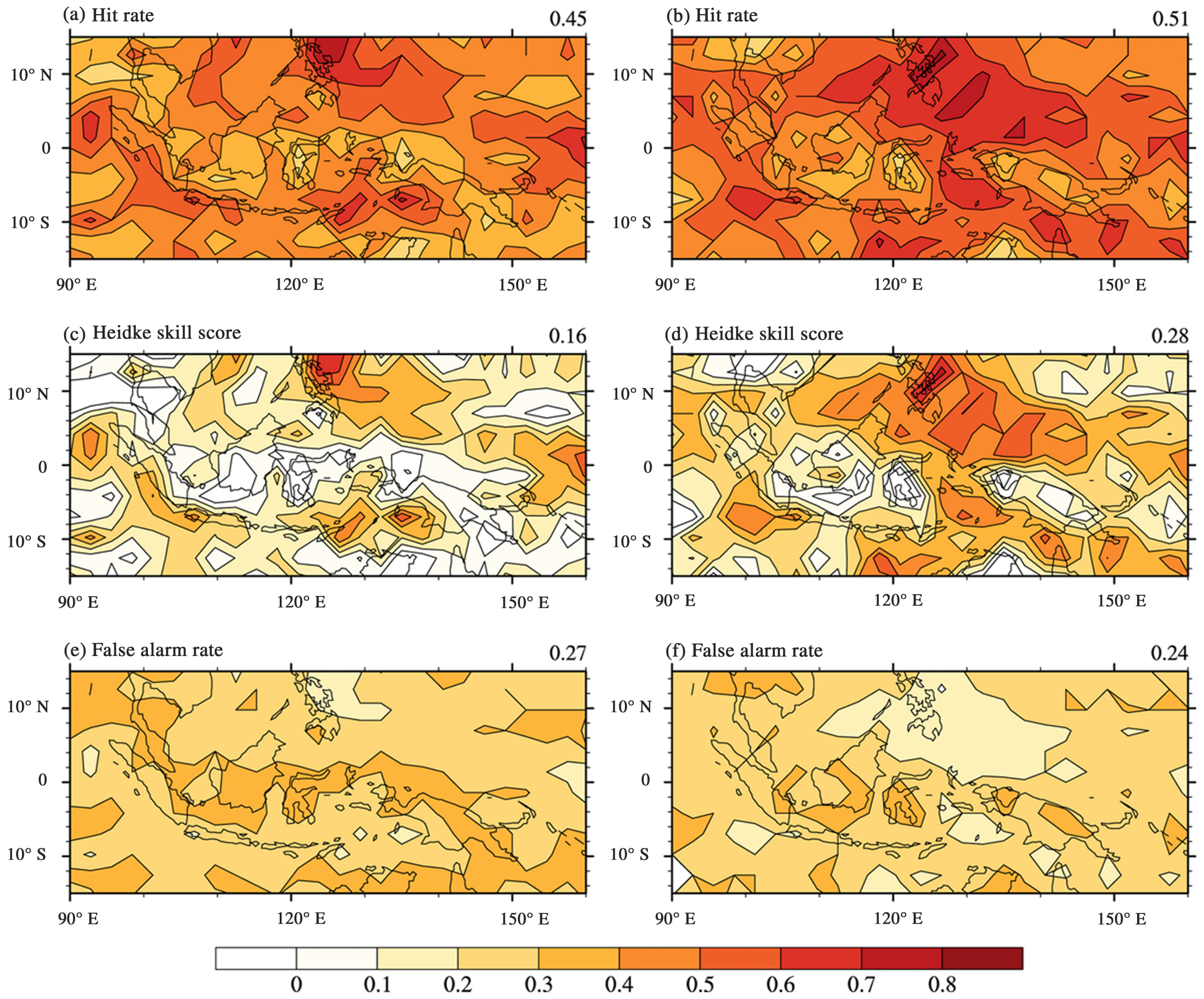

Figure 8. The HR, HSS and FAR of precipitation: (a), (b), (c) for hindcasted and (d), (e), (f) for corrected precipitation.

to the SLR. In addition, all steps apply cross-validation to avoid the over-fitting problem. In previous studies, the prediction results of the CGCM were corrected by producing indices of the arbitrary pattern capable of explaining the largest variances using principal component analysis or creating indices using an area-averaged value in regions that have a high correlation with the correction target statistically (Feddersen *et al.*, 1999; Kang *et al.*, 2004). However, although the predictors selected in the previous methods may be statistically reasonable, they cannot explain the relationship between predictor and predictand physically. In contrast, E-SSLR has the advantage of easily interpreting the result because the physical and dynamical relationships between predictor and predictand are clear.

The predictability of DJF precipitation and OLR of the MC region in PNU CGCM, which is corrected via E-SSLR, is evaluated in this study. First, the predictability of DJF precipitation and OLR, which is corrected via the selection and SLR steps with each index (SSLR), is increased except when SOI is used as a predictor, compared with the hindcasted precipitation and OLR. In the case of SOI, the SSLR result shows lower predictability than the Nino 3.4I case, even though ENSO considerably influences the precipitation in the MC region. This is because the correction using indices poorly simulated by CGCM cannot reflect the relationship between the index and predictand sufficiently. In other words, the effect of correction is dependent on the accuracy of the indices simulated by the model. In addition, when Nino 3.4I and EMI are used as predictor, the seasonal predictability of precipitation and OLR in the MC region is markedly increased. In the case of Nino 3.4I, the correlation between the precipitation and OLR in the MC region is the highest, and TCC between the index predicted by the model and the observed index ranged from 0.89 to 0.96, which is higher than that of the other indices. Similarly, the CGCM predicts EMI better than other indices except for Nino 3.4I and is highly correlated with MC region precipitation and OLR. This leads the dominant improvement of predictability of precipitation and OLR in the region where EMI and variables have high TCC value. This indicates that effect of correction will be increased when the predictor is more influential to the target variable

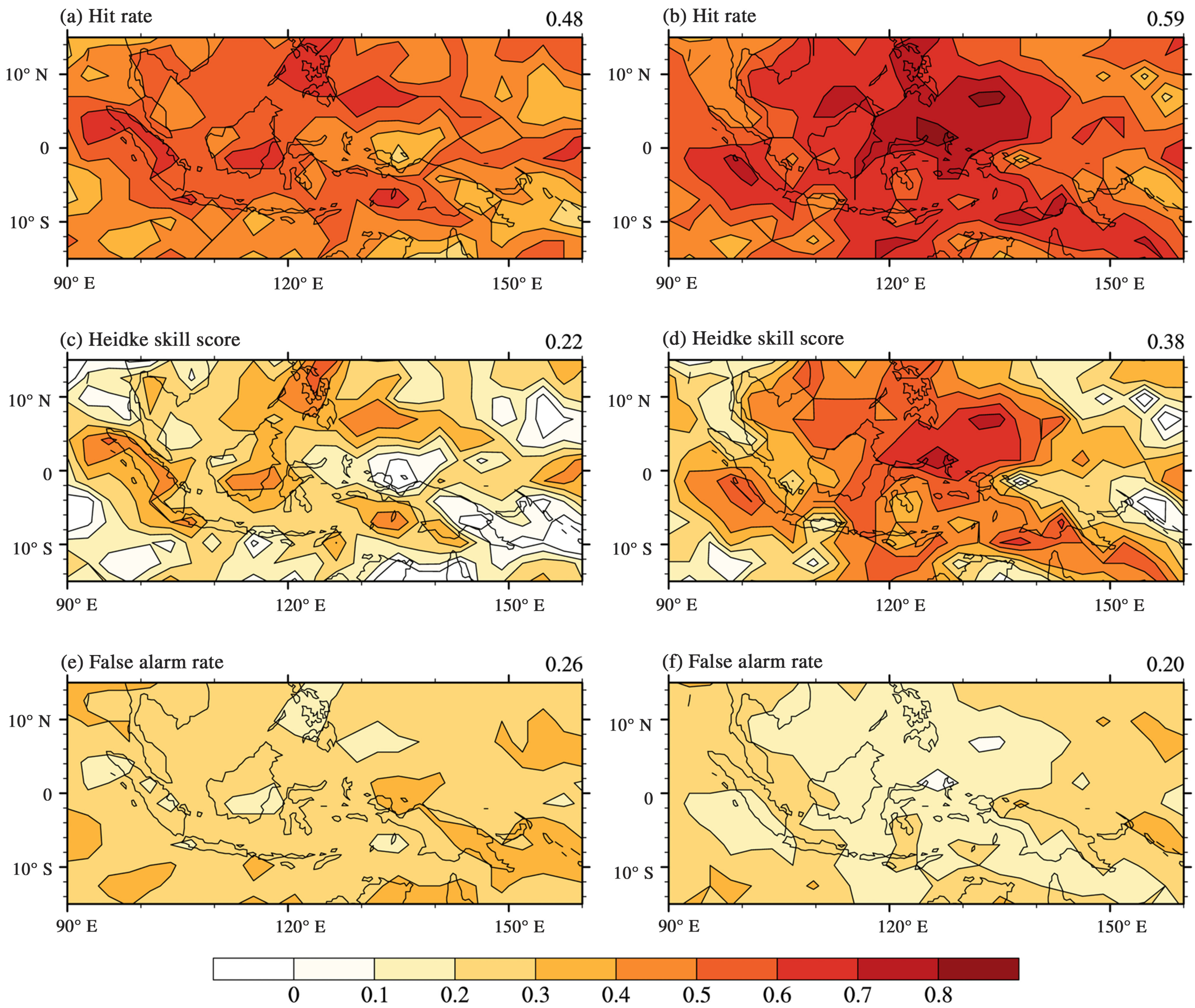

Figure 9. The HR, HSS and FAR of OLR: (a), (b), (c) for hindcasted and (d), (e), (f) for corrected OLR

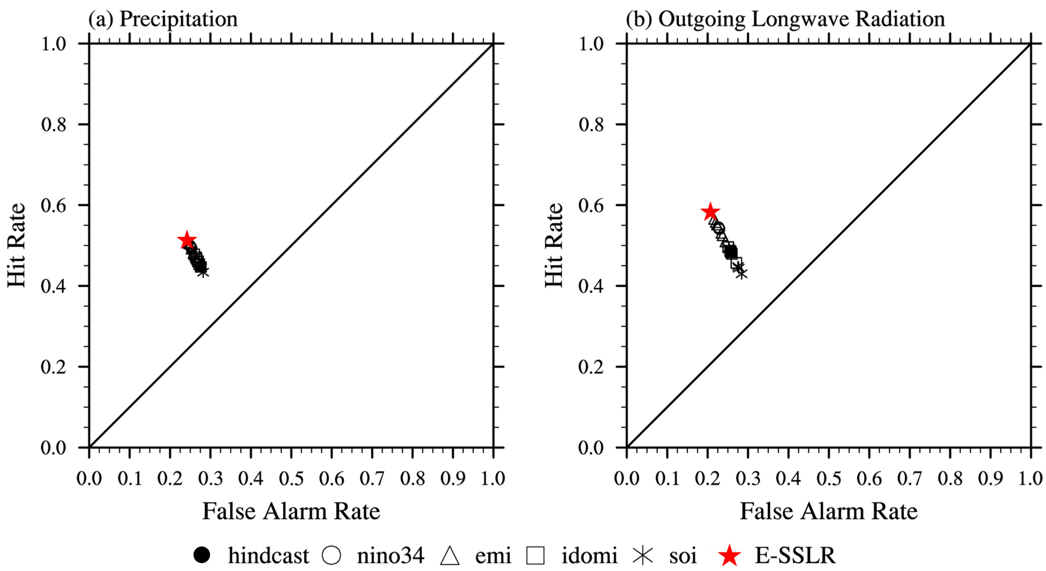

Figure 10. FAR-HR scatterplot of (a) precipitation and (b) OLR.

and more predictable by the model. After SSLR, we reduce the uncertainty of the prediction by performing ensemble (E-SSLR) of the SSLR result, which reflect the influence of each index and improve the seasonal predictability of DJF precipitation and OLR in the MC region.

Acknowledgements

This work was carried out with the support from 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. PJ009953, Republic of Korea.

References

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie PP, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P, Nelkin E. 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.* **4**(6): 1147–1167.

Ahn JB, Kim HJ. 2014. Improvement of 1-month lead predictability of the wintertime AO using a realistically varying solar constant for a CGCM. *Meteorol. Appl.* **21**: 415–418, doi: 10.1002/met.1372.

Ahn JB, Lee JL. 2001. Numerical study on the role of sea-ice using ocean general circulation model. *J. Korean Soc. Oceanogr.* **6**: 225–233.

Ahn JB, Lee JL, Im ES. 2012. The reproducibility of surface air temperature over South Korea using dynamical downscaling and statistical correction. *J. Meteorol. Soc. Jpn.* **90**: 493–507, doi: 10.2151/jmsj.2012-404.

Barnston AG. 1994. Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Clim.* **7**: 1513–1564.

Barnston AG, Thiao W, Kumar V. 1996. Long-lead forecasts of seasonal precipitation in Africa using CCA. *Mon. Weather Rev.* **11**: 506–520.

Behera SK, Krishnan R, Yamagata T. 1999. Unusual ocean–atmosphere conditions in the tropical Indian Ocean during 1994. *Geophys. Res. Lett.* **26**: 3001–3004.

Chang CP, Wang Z, Ju J, Li T. 2004. On the relationship between western maritime continent monsoon rainfall and ENSO during northern winter. *J. Clim.* **17**: 665–672.

Feddersen H, Navarra A, Ward MN. 1999. Reduction of model systematic error by statistical correction for dynamic seasonal prediction. *J. Clim.* **12**: 1974–1989.

Glahn HR, Lowry DA. 1972. The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteorol.* **11**: 1203–1211.

Huang XY. 2000. Variational analysis using spatial filters. *Mon. Weather Rev.* **128**: 2588–2600.

Hurrell J, Hack J, Boville B, Williamson D, Kiehl J. 1998. The dynamical simulation of the NCAR community climate model version 3 (CCM3). *J. Clim.* **11**: 1207–1236.

Jeong HI, Ahn JB. 2007. Long-term predictability for El Nino/La Nina using PNU/CME CGCM. *J. Korean Soc. Oceanogr.* **12**: 170–177.

Jin EK, Kinter JL, Wang B, Park CK, Kang IS, Kirtman BP, Kug JS, Kumar A, Luo JJ, Schemm J, Shukla J, Yamagata T. 2008. Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Clim. Dyn.* **31**: 647–664.

Kanamitsu M, Ebisuzaki W, Woolen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL. 2002. NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorol. Soc.* **83**: 1631–1643, doi: 10.1175/BAMS-83-11-1631.

Kang IS, Lee JY, Park CK. 2004. Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. *J. Clim.* **17**: 834–844.

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

Kug JS, Lee JY, Kang IS. 2008. 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.

Lau KM, Chan PH. 1983. Short-term climate variability and atmospheric teleconnections from satellite-observed outgoing longwave radiation. Part I: simultaneous relationships. *J. Atmos. Sci.* **40**: 2735–2750.

McBride JL, Haylock MR, Nicholls N. 2003. Relationships between the maritime continent heat source and the El Nino-Southern Oscillation phenomenon. *J. Clim.* **16**: 2905–2914.

Michaelsen J. 1987. Cross-validation in statistical climate forecast models. *J. Clim. Appl. Meteorol.* **26**(11): 1589–1600.

Neale R, Slingo J. 2003. The Maritime Continent and its role in the global climate: a GCM study. *J. Clim.* **16**: 834–848.

Pacanowski RC, Griffies SM. 1999. The MOM-3 manual. Technical Report 4, NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ.

Qian JH. 2008. Why precipitation is mostly concentrated over islands in the Maritime Continent. *J. Atmos. Sci.* **65**(4): 1428–1441, doi: 10.1175/2007JAS2422.1.

Ramage CS. 1968. Role of a tropical "Maritime Continent" in the atmospheric circulation. *Mon. Weather Rev.* **96**(6): 365–370.

Saji NH, Goswami BN, Vinayachandran PN, Yamagata T. 1999. A dipole mode in the tropical Indian Ocean. *Nature* **401**: 360–363.

Schiemann R, Demory ME, Mizielinski MS, Roberts MJ, Shaffrey LC, Strachan J, Vidale PL. 2014. The sensitivity of the tropical circulation and Maritime Continent precipitation to climate model resolution. *Clim. Dyn.* **42**(9–10): 2455–2468.

Sun J, Ahn JB. 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.

Sun J, Ahn JB. 2014. Dynamical seasonal predictability of the Arctic Oscillation using a CGCM. *Int. J. Climatol.*, doi: 10.1002/joc.4060.

Sun J, Wang H, Yuan W. 2009. A possible mechanism for the co-variability of the boreal spring Antarctic Oscillation and the Yangtze River valley summer rainfall. *Int. J. Climatol.* **29**: 1276–1284, doi: 10.1002/joc.1773.

Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**: 7183–7192.

Wang B, Lee JY, Kang IS, Shukla J, Kug JS, Kumar A, Schemm J, Luo JJ, Yamagata T, Park CK. 2008. How accurately do coupled climate models predict the leading modes of Asian-Australian monsoon interannual variability? *Clim. Dyn.* **30**: 605–619, doi: 10.1007/s00382-007-0310-5.

Webster PJ, Moore AM, Loschnigg JP, Leben RR. 1999. Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–1998. *Nature* **401**: 356–360.

Wilks DS. 1995. *Statistical Methods in the Atmosphere Science*. Academic Press: San Diego, CA, 220 pp.