Asia-Pac. J. Atmos. Sci., 50(3), 307-326, 2014
DOI:10.1007/s13143-014-0019-4
pISSN 1976-7633 / eISSN 1976-7951

Validation of the Experimental Hindcasts Produced by the GloSea4 Seasonal Prediction System

Myong-In Lee[1], Hyun-Suk Kang[2], Daehyun Kim[3], Dongmin Kim[1], Hyerim Kim[1], and Daehyun Kang[1]

[1]*School of Urban and Environmental Engineering, UNIST, Ulsan, Korea*
[2]*Climate Research Division, National Institute of Meteorological Research, Jeju, Korea*
[3]*Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, U. S. A.*

(Manuscript received 11 November 2013; accepted 18 November 2013)

Abstract: Using 14 year (1996-2009) ensemble hindcast runs produced with the Global Seasonal Forecasting System version 4 (GloSea4), this study evaluates the spatial and temporal structure of the hindcast climatology and the prediction skill of major climate variability. A special focus is on the fidelity of the system to reproduce and to forecast phenomena that are closely related to the East Asian climate. Overall the GloSea4 system exhibits realistic representations of the basic climate even though a few model deficiencies are identified in the sea surface temperature and precipitation. In particular, the capability of GloSea4 to capture the seasonal migration of rain belt associated with Changma implies a good potential for the Asian summer monsoon prediction. It is found that GloSea4 is as skillful as other state-of-the-art seasonal prediction systems in forecasting climate variability including the El-Niño/southern oscillation (ENSO), the East Asian summer monsoon, the Arctic Oscillation (AO), and the Madden-Julian Oscillation (MJO). The results presented in this study will provide benchmark evaluation for next seasonal prediction systems to be developed at the Korea Meteorological Administration.

Key words: Seasonal prediction, GloSea4, ENSO, MJO, Asian monsoon, AO

1. Introduction

Global prediction systems have been developed for operational medium-range weather forecasts and seasonal prediction (e.g., Molteni *et al*., 1996; Kusunoki *et al*., 2001; Saha *et al*., 2006, 2013; Arribas *et al*., 2011, hereafter A11; Kim *et al*., 2012). These are now being advanced into systems based on fully-coupled climate system models that include comprehensive dynamics and physics of atmosphere, land surface, ocean, and sea ice interactions. Recent studies suggest that the prediction skill of these systems still has much room for further improvement. For example, Wang *et al*. (2009) pointed out that individual prediction by the coupled models exhibit significant biases and limitations over the tropics and East Asian monsoon region, and suggested that the multi-model ensemble (MME) method was valuable to reduce systematic biases and to quantify the forecast uncertainty. Jin and Kinter (2009) discussed the limits of El Niño and the Southern Oscillation (ENSO) prediction by a coupled model, and showed a clear seasonal dependency in the prediction skill known as the "spring predictability barrier" (Webster and Yang 1992; Webster 1995; Torrence and Webster 1998; Jin *et al*., 2008). A more recent assessment of the seasonal prediction skill for the Northern Hemisphere winter using long-term hindcasts from the European Center for Medium-range Weather Forecasts (ECMWF) and the National Center for Environmental Prediction (NCEP) systems showed somewhat improved skill in the ENSO prediction and teleconnection patterns (Kim *et al*., 2012). However, both systems have difficulty in the winter temperature prediction in the mid-latitudes over the United States and northern Europe, presumably due to the low predictability in other climate variability such as the North Atlantic Ocean (NAO).

It has been suggested that both the global model and the initialization of the prediction system require further improvements. The horizontal and vertical resolution of the prediction systems has been continuously increased. The horizontal resolution of the ECMWF Integrated Forecast System (IFS; Gregory *et al*., 2000; Bechtold *et al*., 2008) model has been increased from TL159 (~125 km) in System 3 to TL255 in System4 (~80 km). Recently, Met Office has also increased the resolution of the seasonal prediction system into N216L85 for atmosphere, which is equivalent to 40 km horizontal resolution with 85 vertical levels (Maidens *et al*., 2013). NCEP is now testing a T382 (~30 km) resolution version of the Climate Forecasting System (CFS) for the dynamical hurricane prediction (Schemm *et al*., 2011). There have been many studies that emphasize the importance of the physical parameterizations in seasonal prediction. Buizza *et al*. (1999) improved the medium-range prediction skill by stochastically perturbing tendencies calculated from the physical parameterizations in the system. Studies of Shin *et al*. (2003), Byun and Hong (2007) and Ham and Hong (2013) suggest an important role of the cumulus parameterization scheme in the seasonal prediction skill. In particular, Bechtold *et al*. (2008) and Park *et al*. (2010) showed that the prediction of the tropical subseasonal vari-

Corresponding Author: Myong-In Lee, School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, UNIST-gil 50, 100 Banyeon-ri, Ulju-gun, Ulsan 689-798, Korea.
E-mail: milee@unist.ac.kr

ability could be improved by improving cumulus parameterization. Also there are an increasing number of studies focusing on the initialization process of the coupled prediction system. For example, Koster *et al*. (2010) suggested that the land surface initialization may contribute, at least over the North America, to the prediction skill of precipitation and surface temperature. A new method of using empirical singular vectors in perturbing initial conditions in the ensemble seasonal prediction was developed by Kug *et al*. (2010). Ham and Rienecker (2012) showed an improvement of the ENSO prediction by improving ensemble generation method from their 20 year-long hindcast experiments. Recent active satellite observations in Arctic sea ice concentration are also making it possible to initialize the sea ice model.

A systematic validation of the prediction system based on long-term, historical reforecast runs (i.e., hindcasts) is one of the basic efforts needed to provide a useful guidance for its further improvement. The Korea Meteorological Administration (KMA) has been collaborating with the United Kingdom Met Office (UKMO) for the past years to develop comprehensive and reliable operational forecast systems that cover from the short-range weather to seasonal time scale, and even to the climate projection of multiple decades. For seasonal prediction, the Global Seasonal Forecasting System version 4 (GloSea4) developed by the Met Office Hadley Center (A11) has been tested, which is a state-of-the-art prediction system based on a fully coupled climate system model. A series of hindcasts for the past 14 years (1996-2009) have been conducted at the local supercomputer at KMA (Kang *et al*., 2011), and this is the KMA's first hindcast experiment using a fully coupled climate prediction system for a routine operation.

The main objective of this study is to evaluate the basic performance of the GloSea4 hindcasts. Evaluation is performed specifically focusing on two aspects: quality of hindcast climatology and prediction skill of major climate variability. Examination of the hindcast climatology is to identify the systematic bias of the model thereby providing a useful guidance for further model development. Under the first objective, we examine the global energy balance, and patterns of global sea surface temperature (SST) and precipitation. Specifically, the model depictions in the East Asian summer monsoon and the diurnal cycle of warm season precipitation are examined, which are useful diagnostics for the moist physics parameterization of the prediction system. We aim at guiding practical use of the prediction system by assessing the prediction skill of major climate variability. In this study, we examine the prediction skill of ENSO, the Arctic Oscillation (AO), and the Madden-Julian Oscillation (MJO), which are known to affect the weather and climate variability in East Asia significantly, as well as the East Asian summer monsoon itself. We have to mention that the current work is supplementary to the study of A11, in which some of the evaluations in this work were performed, with a primary focus on the fidelity of the system to reproduce and to forecast phenomena that are closely related to the East Asian climate. Also note that a few changes have been made in the GloSea4 prediction system since A11 (for example, the inclusion of sea ice initialization), so that the dataset we are using is different from that used in A11. The possible impacts of these changes in the performance of the system will be discussed in the text too.

Sections are followed as: Section 2 provides a brief description of GloSea4 hindcast experiments; Section 3 the examination of the hindcast climatology; Section 4 the prediction skill for major climate variability (i.e., ENSO, summer monsoon, AO, and MJO). Section 5 summarizes the results and provides major conclusions.

2. Data and method

a. GloSea4 system

GloSea4 is the fourth version of the global ensemble seasonal prediction system developed at UKMO and became operational since September 2009. The configuration of GloSea4 used in this study is basically same as that used in A11, with a few changes. Only a minimal description to the system and important changes from A11 is provided here for the discussion purpose. Interested readers are referred to A11 for more details.

The global model embedded in GloSea4 is the Hadley Center Global Environment Model version 3 model (HadGEM3), which is a fully-coupled global climate model with atmosphere, land surface, ocean, and sea-ice components (Hewitt *et al*., 2011). Atmospheric model is based on the Met Office Unified Model (UM; Davies *et al*., 2005), with the land surface model of Met Office Surface Exchange Scheme (MOSES; Essery *et al*., 2003). Ocean model is based on Nucleus for European Modeling of the Ocean (NEMO; Madec, 2008), coupled with the Los Alamos sea ice model (CICE; Hunke and Lipscomb, 2010). Each component model is connected though the OASIS coupler developed by Valcke (2006).

Since the study of A11, the version of HadGEM3 has been updated from r1.1 to r4.0 with the upgrade of Global Atmosphere version 2.0 (GA2.0) atmospheric model. Vertical resolution of the atmospheric model has been increased from 38 to 85 levels. The vertical resolution of the ocean model has been also increased from 42 to 75 levels, whereas the horizontal resolution of the ocean model remains same (1/3^{o} between 20^{o}S and 20^{o}N and 1^{o} for the rest of the domain). Note that the most recent operational version at UKMO and KMA is the GloSea version 5 (GloSea5, Maidens *et al*., 2013), which has a newer version of the atmospheric model (GA3.0) and with increased horizontal resolutions of the component models: GA3.0 in N216L85 (~60 km in mid-latitudes), NEMO and CICE in ORCA025L75 (1/4^{o} resolution in the tropics of 20^{o}S and 20^{o}N). Hindcast data from this newer operational system are currently being produced at KMA, which are not available yet for a complete analysis.

Table 1. Datasets used for GloSea4 validation.

Validation Data	Acronym	Resolution	Variables	Validation Period	Frequency	Reference
Clouds and the Earth's Radiant Energy System	CERES	$1.0^o \times 1.0^o$	Radiation	2000-present	Monthly	Wielicki *et al.* (1996)
Global Precipitation Climatology Project	GPCP	$1.0^o \times 1.0^o$	Precipitation	1996-2009	Monthly	Adler *et al.* (2003)
CPC Merged Analysis of Precipitation	CMAP	$2.5^o \times 2.5^o$	Precipitation	1996-2009	Pentad	Spencer (1993)
Modern Era-Retrospective Analysisfor Research and Applications	MERRA	$1/2^o \times 2/3^o$	SLP, Wind, Surface Temperature	1996-2009	Monthly, Daily	Bosilovich (2008)
Optimum Interpolation Sea Surface Temperature	OISST	$0.25^o \times 0.25^o$	SST	1996-2009	Daily	Reynolds *et al.* (2002)
Advanced 'Very High Resolution Radiometer	AVHRR	$2.5^o \times 2.5^o$	OLR	1996-2009	Daily	Hastings *et al.* (1992)
ECMWF Re-Analysis	ERA-interim	$1.5^o \times 1.5^o$	Temperature, Humidity, Wind, Geopotential Height, Surface Temperature	1996-2009	Daily	Dee *et al.* (2011)
Tropical Rainfall Measuring Mission	TRMM	$0.25^o \times 0.25^o$	Precipitation	1998-2009	3-Hourly	Huffman (2007)

b. Hindcast experiment

Long-term hindcast experiments using GloSea4 have been conducted for 14 years (1996-2009) in this study. Each hindcast run was initialized using the ERA-interim reanalysis for atmosphere and land surface (Dee *et al*., 2011). An anomaly initialization technique is used for land surface, where the anomalies of initialized variables (e.g., soil moisture) are calculated first by subtracting the ERA-interim climatology and then added to the HadGEM3 model climatology. This is to avoid the possible drift of the forecast field from the basic state of ERA-interim to that of HadGEM3. It is noted that the prescription of the initial states for atmosphere and land in the hindcasts are different from that in the forecast mode of GloSea4. For the real-time forecasts, the atmosphere and land surface conditions are initialized using the Met Office operational numerical weather prediction (NWP) analysis. Ocean state was initialized using the data from the Met Office Ocean Data Assimilation (ODA) system in the hindcasts, which is identical to the method used in the real-time forecast except that the Met Office NWP fluxes are replaced by ERA-interim fluxes. The sea ice initialization scheme has been changed since A11. Seasonally-varying model climatology was prescribed in A11, whereas sea ice was initialized by the ocean data assimilation in the hindcasts used in the current study. Obviously, the new sea ice initialization method is a more realistic one, and therefore it is reasonable to expect enhancement of forecast skill of high latitude climate variability.

Hindcast runs were started at fixed calendar dates - 1st, 9th, 17th and 25th - in each month, and integrated up to 7 months. Ensemble runs in GloSea4 do not explicitly represent the uncertainties in the initial conditions (A11). Instead the model uncertainties are represented by the use of the stochastic kinetic energy backscatter scheme version 2 (SKEB2; Shutts, 2005; Tennant *et al*., 2011). This yields three ensemble runs for each starting date of the hindcasts.

In the actual operation of GloSea4 at Met Office, a set of hindcast were concurrently produced in real time with the forecast. This is different from the conventional procedure taken at other centers such as NCEP where all hindcast simulations are completed before running any real-time forecast. As discussed in A11, this approach has an advantage in that one does not need to repeat complete and expensive hindcast runs at frequent changes in the global model and initialization methods. This is practically useful for dealing with the huge demand in computing resource in ensemble seasonal prediction. However, a relatively short hindcast period of 1996-2009 in this study is limiting a robust evaluation to the historical forecast skill. In addition, the predictability change in the mean state and the reproduction capability by the prediction system cannot be addressed accordingly.

c. Validation methods and data

In this study, we define hindcast climatology as a function of the initialization month and the forecast lead time. Therefore, 168 hindcasts (3 ensemble members per week by 4 weeks by 14 years of the 1996-2009 period) are averaged to obtain a hindcast climatology initialized from a specific month. This hindcast climatology has been analyzed to identify the systematic bias of the global model and the seasonal prediction system. Various satellite observations and reanalysis datasets were used for the validation. Table 1 shows the validation datasets and their period, data frequency, and the references. For evaluations of the forecast climatology, we mostly analyzed the 3 month-lead hindcast. During this forecast lead time, we expect that the influences of the initial states are gradually decaying while the model deficiencies are gradually contributing more to the forecast errors. When evaluating forecast skill of climate variability, the hindcast climatology for the

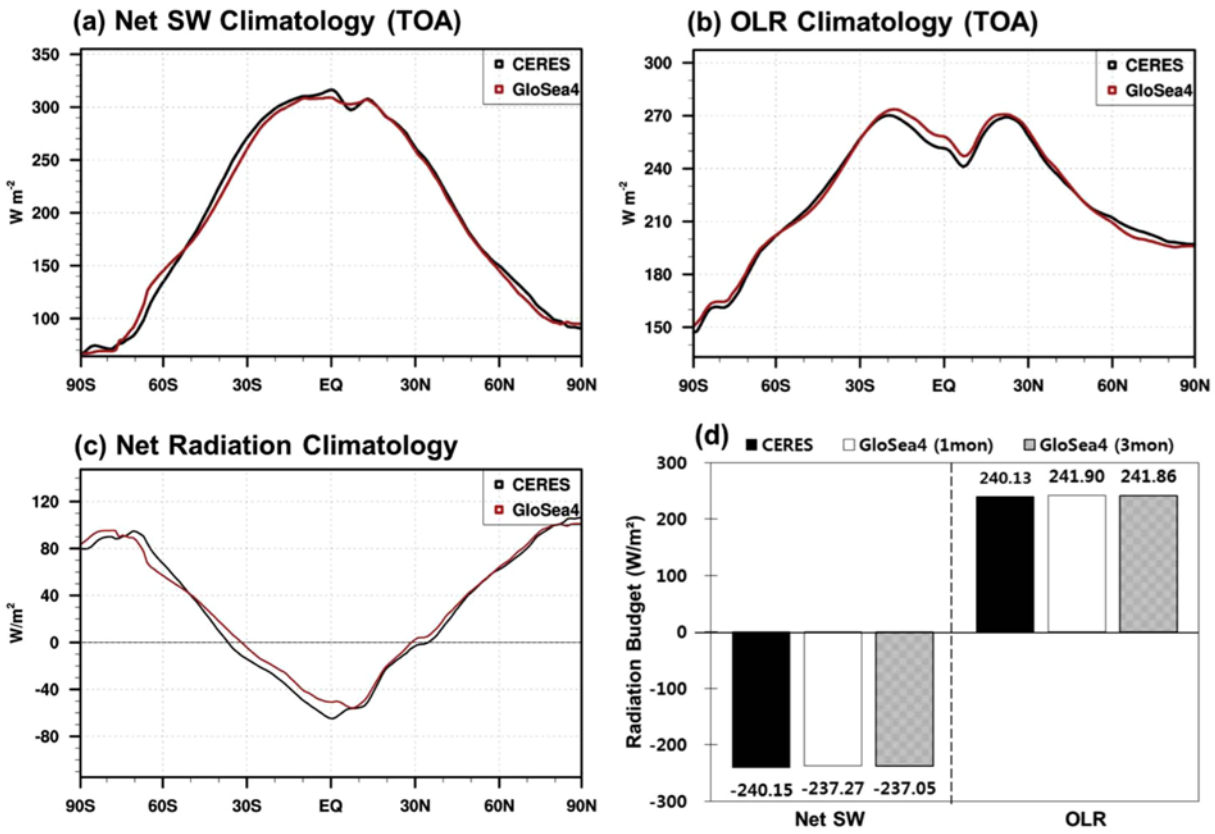

Fig. 1. Comparisons of the zonal-mean annual-mean (a) net shortwave (SW) radiation, (b) OLR, and (c) net radiation ((a) minus (b)) at TOA between CERES and GloSea4. The climatological mean is defined for the period of 2000-2009 when CERES is available. (d) is the global mean radiative fluxes from CERES and GloSea4 with 1 month lead and 3 month lead hindcasts, respectively. Negative value is defined as the downward flux, and the unit is W m^{-2}.

given forecast lead time is removed in all the prediction fields, and these anomalies were used. The validation methods and approaches in this study rely on general and popular choices taken in the previous literatures for a fair comparison. The specific methods will be briefly discussed appropriately in the following Sections.

3. Structures of climatology

It is important that the improvement in the accuracy of the seasonal prediction should be attained in robust and reliable structure of the simulated climatology. In this section, we evaluate the spatial and temporal structures of the hindcast climatology produced by GloSea4.

a. Energy balance at TOA

The radiation budget at the top of atmosphere (TOA) is an important aspect of a coupled global climate system. The net TOA radiation flux (the sum of the OLR and the net incoming shortwave (SW) radiation) is supposed to be close to zero when a long-term average is taken. In practice, however, this is hardly zero in the seasonal prediction; imbalance between initial condition and model climatology, and deficiencies in the model could introduce a drift of the mean state, which would produce a non-zero net TOA radiation flux. In Fig. 1, zonal mean patterns of the net downward SW radiation and OLR are compared between CERES observations and GloSea4. The radiation fluxes of GloSea4 are remarkably good and quite comparable to those from CERES. When the global mean values are compared (Fig. 1d), the net SW radiation is slightly underestimated with a difference around 3 W m^{-2}, whereas OLR is slightly overestimated with a difference less than 2 W m^{-2}. In the net radiation, about 4.6 W m^{-2} residual radiative flux goes out of the system in GloSea4, while the residual radiation in CERES is very close to zero (–0.02 W m^{-2}), exhibiting a good energy balance at TOA from these 9-year observations. The energy imbalance at TOA seems a common problem in many climate models. Even the reanalysis products experience this kind of energy imbalance, often with a greater magnitude of bias. A larger planetary albedo in GloSea4 than CERES (not shown) seems to be partly responsible for the underestimation of net SW radiation. The energy imbalance at TOA in GloSea4 is slightly enhanced as the forecast lead time increases from +4.6 W m^{-2} in the 1-month lead hindcasts to +4.8 W m^{-2} in the 3-month lead hindcasts. The most of the energy imbalance at TOA is compensated by more upward energy flux at the surface in GloSea4 (+4.4 W m^{-2} in the 3-

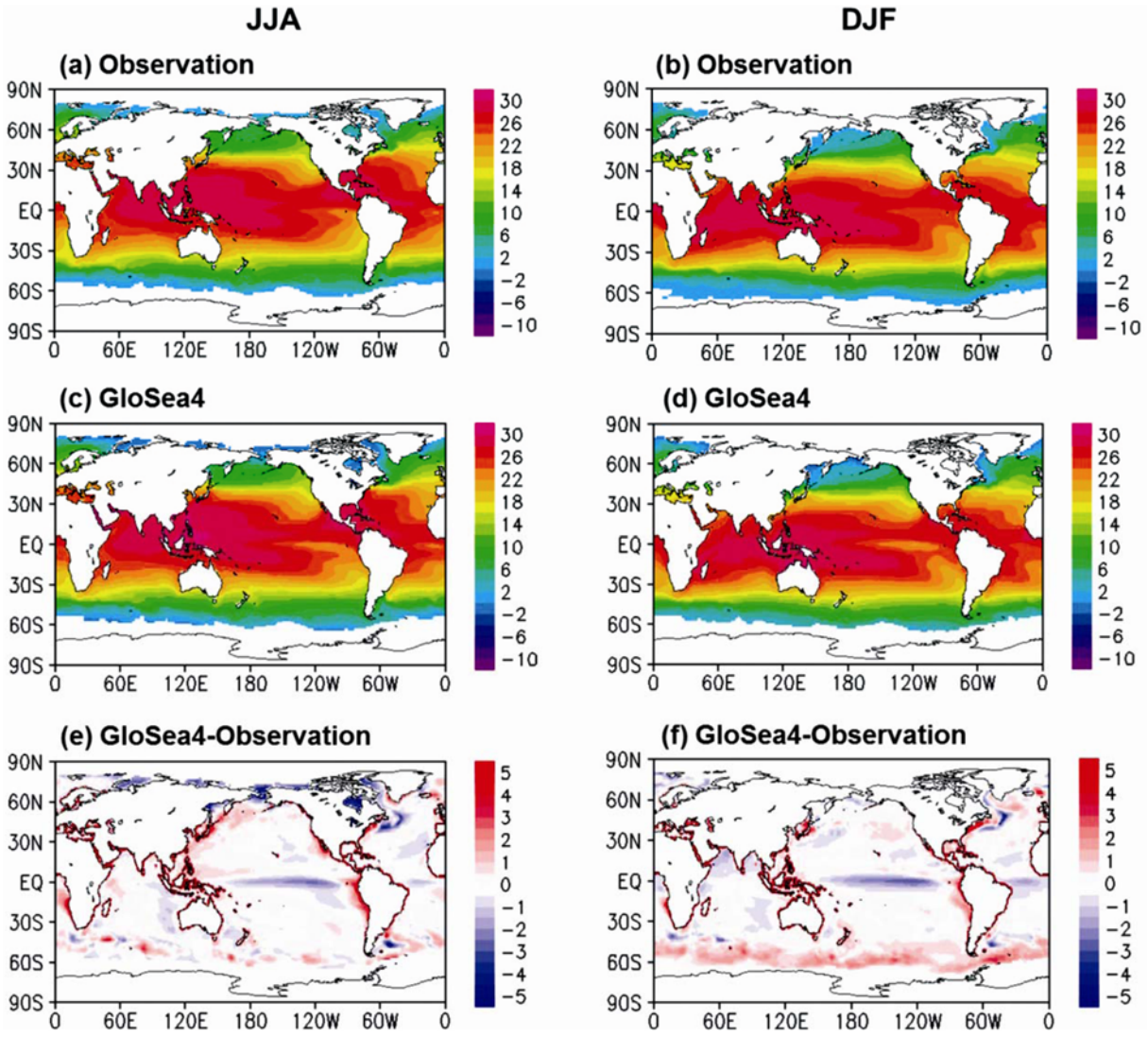

Fig. 2. Seasonal-mean SST in JJA (left) and DJF (right) from OISST observation (top) and GloSea4 hindcasts with 3 month lead time (middle). Bottom panels indicate the model bias (model minus observation). The unit is °C.

month lead hindcasts). During initial 4 months, atmosphere in the GloSea4 system looses about 0.4 W m^{-2}, which seems to be contributing to the overall cold bias in the atmospheric temperature in a comparison with the ERA-Interim reanalysis (not shown).

b. Seasonal means

Accuracy of the SST prediction is one of the most important factors that define the quality of the seasonal prediction system, as a pronounced portion of the interannual variability is explained with the tropical SST variation. Figures 2a-d compare the JJA and DJF SST patterns from the OISST observations with those constructed using 3-month lead data from GloSea4 hindcasts. GloSea4 represents the observed SST patterns reasonably well in both seasons, such as the Indo-Pacific warm pool region and the cold pool region in the equatorial eastern Pacific. The sharp gradient of SST from subtropics to high latitudes is also reasonably resolved in GloSea4. As shown in Figs. 2e, f, a systematic error in the SST prediction is also noticeable, particularly in the equatorial eastern and central Pacific where the observed cold SST region overestimated. This bias is often referred to as the "cold tongue" problem, which is common in many ocean-atmosphere coupled climate models, and the cause of this bias is not clearly understood yet (e.g., Mechoso *et al*., 1995; Yu and Mechoso, 1999; Latif *et al*., 2001; Davey *et al*., 2002; Misra *et al*., 2008). In Fig. 3, the bias pattern of SST in the equatorial Pacific region is compared with that of the cloud radiative forcing (CRF), which is defined as the difference between radiative fluxes between all-sky condition and clear-sky condition. It is noted that the CRF bias pattern is coinciding well with the SST bias pattern, which suggests they might be related to each other. When the SW and longwave (LW)

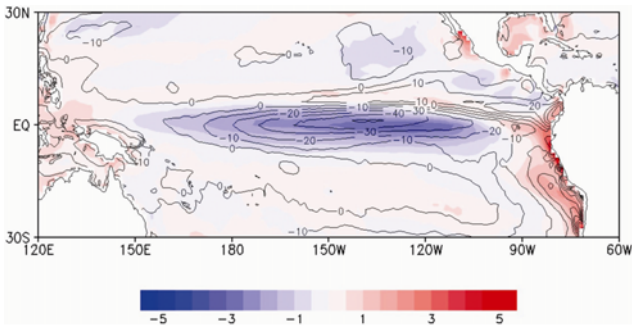

Fig. 3. The model biases in the net cloud radiative forcing (contour, unit: W m^{-2}) and SST (shaded, K). The model bias is estimated against the CERES observations for the cloud radiative forcing, and the OISST for SST, respectively.

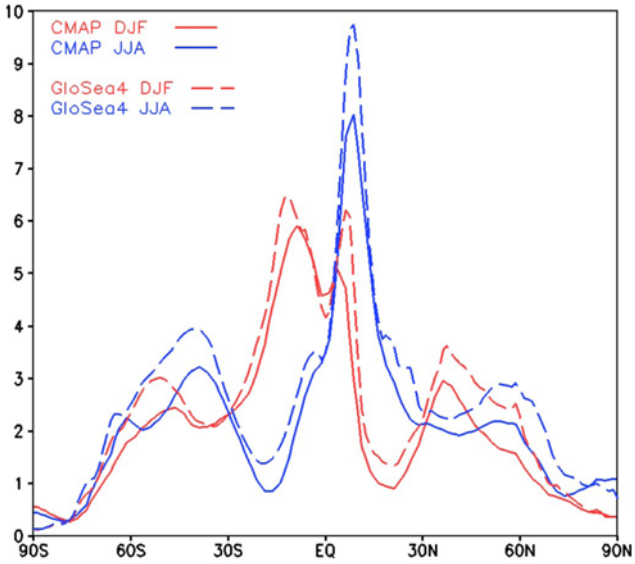

Fig. 4. Comparison of the zonal-mean precipitation between CMAP observations and GloSea4 hindcast climatology with 3-month lead time. Observations are indicated in solid lines and GloSea4 are in dashed lines. The unit is mm d^{-1}.

radiation components of CRF were separately analyzed in the biased region, it was found that the SW cloud effect is overestimated and responsible for the reduction of solar radiation coming into the equatorial tropical ocean. The cold tongue bias as well as the bias in other oceanic areas does not depend much on season.

Zonal-mean patterns of precipitation in JJA and DJF seasons are compared in Fig. 4. Overall, the model overpredicts the zonal-mean precipitation, regardless of season. Particularly the time-mean precipitation is overpredicted in boreal summer in the ITCZ region. Zonal-mean evaporation and the meridional moisture transport are also larger than the observed in the simulation (not shown), reflecting a too excessive water cycle in this model.

Figure 5 compares the horizontal distribution of precipitation in JJA and DJF. In observations from CMAP (Figs. 5a, b), the mean precipitation pattern is characterized by the peaks over the western Pacific, south China sea, Bay of Bengal, south eastern equatorial Indian Ocean, as well as the zonally-elongated rainbands (intertropical convergence zone, ITCZ) over the Pacific and Atlantic ocean. The Glosea4 hindcast climatology (Figs. 5b, d) captures the observed wet regions reasonably well, although wet and dry biases are notable in some regions (Figs. 5e, f). In JJA (Fig. 5e), the simulation tends to overestimate rainfall over the northeast of the Arabian Sea, the northeast of the Bay of Bengal, south of Sahel, western Indian Ocean, and the eastern end of the equatorial Pacific. Wet bias is also found in the Atlantic ITCZ. On the other hand the model simulation lacks precipitation over the western Pacific and eastern Indian Ocean. In DJF (Fig. 5f), the wet bias over the Andes is noticeable, pairing with the wet bias in Himalayas in JJA. The regions of equatorial Africa, the western Indian Ocean, and the central Pacific ITCZ are also the regions of overestimation. The Atlantic ITCZ is shifted northward in GloSea4, causing dry bias in the northeast of the Amazon.

The comparison of DJF and JJA bias patterns indicates systematic biases of the simulated rainfall regardless of season. These include i) a general wet bias over the land area, especially near big mountains, ii) wet bias over the Maritime continent and dry bias over the surrounding oceanic area, iii) Indian Ocean dipole mode-like bias pattern over the Indian Ocean, iv) overestimation of the strength of central-to-eastern Pacific ITCZ, Atlantic ITCZ, and South Pacific convergence zone (SPCZ). It is also noted that the wet biases straddling north and south of the equator and dry bias at the equator in GloSea4 seems not to be totally independent of the cold tongue bias in SST (see Fig. 2). This bias tends to drive a double-ITCZ pattern in the central Pacific (Schneider, 2002; Zhang and Wang, 2006; Lin, 2007; Hwang and Frierson, 2013).

The Taylor diagram is another quantitative measure for the fidelity of precipitation simulation. For validation, we downgraded the horizontal resolution of the observed GPCP data onto a coarse resolution that is equivalent with GloSea4. In comparison with GPCP, the spatial correlation is slightly below 0.9, regardless of season, demonstrating a high capability of reproduction for the time-mean precipitation patterns. On other hand, the spatial variance is underestimated in GloSea4, which signal is getting clearer in the longer forecast lead time.

Although not being presented, we note that the zonal-mean distributions of the simulated CRF are quite realistic in terms of meridional structure as well as the actual magnitude. We further examined the simulated distribution of cloud using the recent CloudSat satellite observation (Li *et al*., 2012; Graeme *et al*., 2008). In Fig. 7, the zonal mean liquid and ice cloud amounts are separately compared. Only the annual mean is examined here due to data availability. As in observations, cloud liquid water simulated by the GloSea4 system is concentrated between 900-700 hPa levels in extratropics, while it extends vertically up to 500 hPa in the tropics because warm temperature bias shifts the freezing level higher than the observed. Compared to observations, GloSea4 in general under-

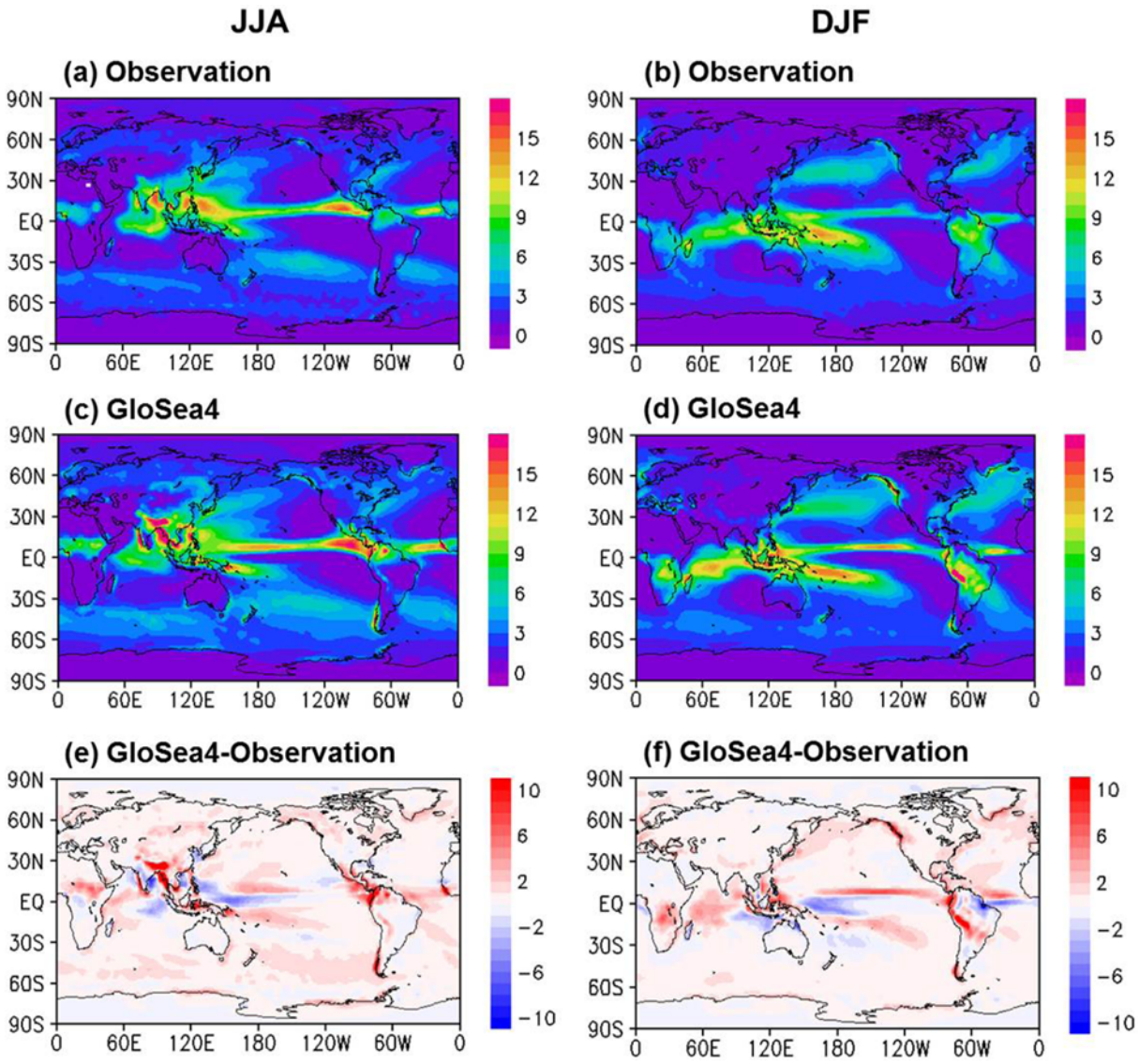

Fig. 5. Horizontal distribution of climatological-mean precipitation in JJA (left) and DJF (right). CMAP observations are in top, GloSea4 with 3 month lead time in middle, and the model biases are in bottom. The unit is mm d^{-1}.

estimates the amount of cloud liquid water over the mid-latitudinal region in both hemispheres. For the ice cloud amount the observation shows a concentration above 500 hPa in the tropics, while it goes down with latitude in both hemispheres. GloSea4 tends to simulate the meridional distribution reasonably well. The model, however, underestimates the amount of cloud ice water in the tropical deep convective region, whilst it excessively overestimates in the extra-tropical area.

c. East Asian monsoon

GloSea4 exhibits a realistic simulation of the summer monsoon in East Asia, as well as its seasonal evolution. Figures 8a, b are the seasonal mean precipitation and the 850 hPa zonal (u-) wind pattern over the South and East Asian domain during monsoon season (May-August) in observations and GloSea4 hindcasts, respectively. Seasonal cycle of precipitation and 850 hPa zonal wind at two longitude bands (70-80°E and 120-130°E) are also shown in Figs. 8c-f. As the GloSea4 hindcast runs are available up to 7 months from the initial time, two hindcast data from the February and August initialized runs were synthesized to examine the complete annual cycle of precipitation and wind. Local maxima of monsoon precipitation are realistically captured by GloSea4, such as the northeast of the Arabian Sea and the Bay of Bengal, and the west of the Philippines. These oceanic regions are the upstream side of the low-level monsoon flow before the elevated land surface, coinciding with the convergence area of 850-mb u-winds, and these aspects are also reasonably captured by GloSea4. Precipitation amount over these regions tend to be overestimated in GloSea4, which is also a common feature in current global

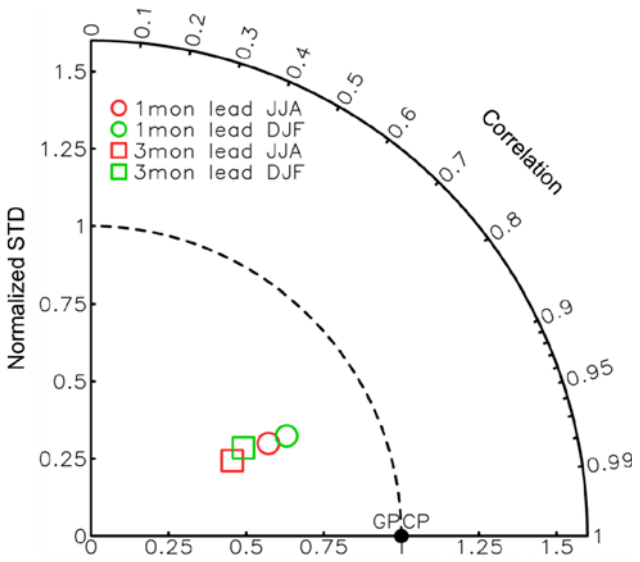

Fig. 6. The Taylor diagram of time-mean precipitation simulated by GloSea4 in comparison with GPCP observations. JJA (June-July-August) denotes the boreal summer precipitation, and DJF (December-January-February) is the boreal winter.

prediction models, particularly at the foothill side of the Himalayas. Another notable deficiency is in the tropical Western Pacific, where GloSea4 underestimates the rainfall amount. The simulation is remarkably good over China, Korea, and Japan where the GloSea4 simulation is able to capture the zonally-elongated rainband associated with Changma. The seasonal march of precipitation in the Indian monsoon (70-80°E) and the East Asian monsoon domains (120-130°E) coincides well between observation and model hindcast climatology (Figs. 8c-f). For example, the northward march of the Indian monsoon rainband during April to July (Figs. 8c, d) is represented successfully well. In the East Asian domain (Figs. 8e, f), the bifurcation of rainband, one in South China Sea and another in the Far East, during June to August is related with the northward movement of the Changma front. GloSea4 exhibits rather fast northward migration of the monsoon rainband, which reaches further north to 40°N. Associated with this, observed precipitation over Korea and Japan exhibits a secondary peak in late summer (September), which is also captured by GloSea4. Overall, the monsoon simulation by GloSea4 is realistic, suggesting a potential to predict the Asian summer monsoon with a high fidelity.

d. Diurnal cycle of precipitation

As one of the fastest cycles resolved in the prediction system, we compared the observed diurnal cycle from the TRMM 3-hourly rain rate data and the GloSea4 simulation

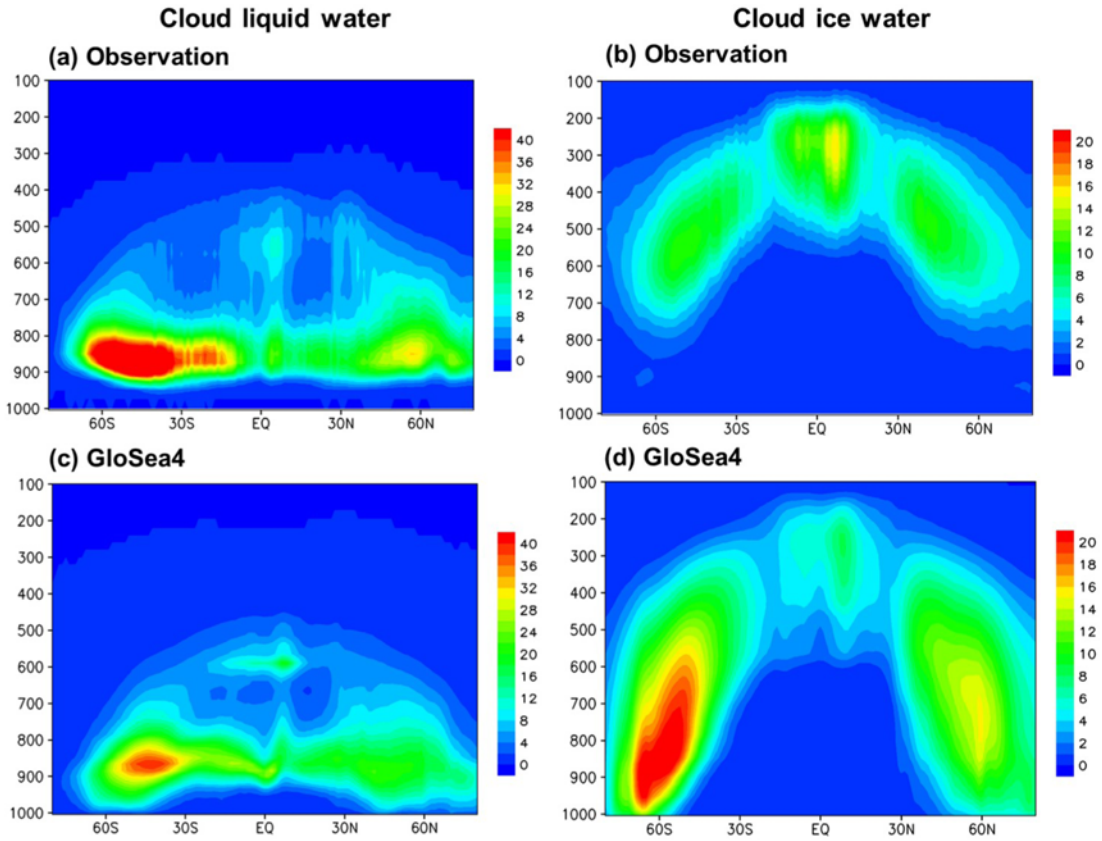

Fig. 7. Comparisons of annual-mean zonal-mean cloud liquid and ice amount from the CloudSat satellite observations (top) GloSea4 hindcasts with 3 month forecast lead time (middle). Bottom panels show the model bias. The left column shows the cloud liquid water, and the right column shows the cloud ice water. The unit is g m^{-3}.

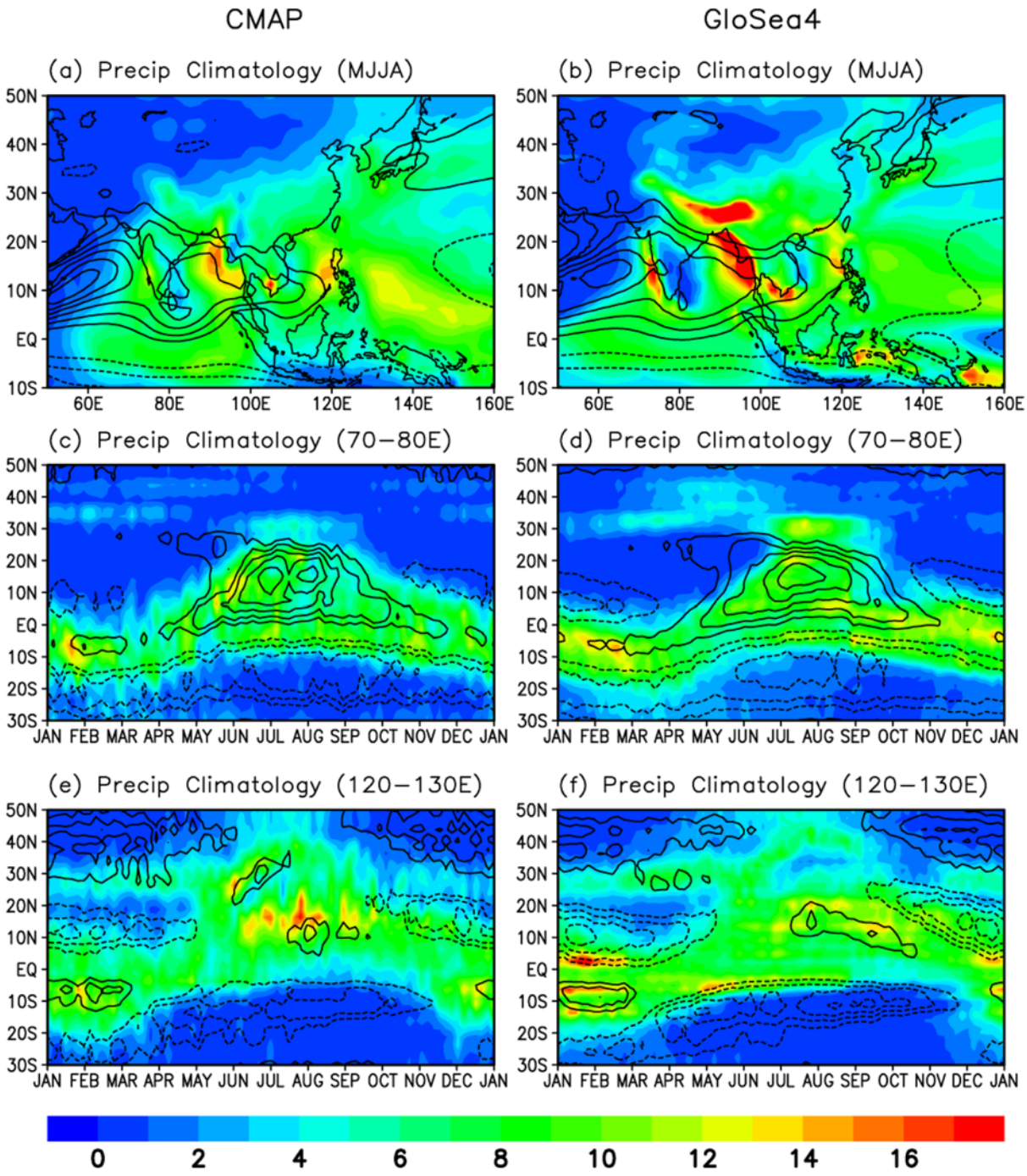

Fig. 8. Climatological-mean precipitation (shaded) and zonal wind at 850 hPa (contour, above 3 m s^{-1}) from (a) CMAP and ERA-interim, and (b) GloSea4 during May to August averaged over 14 years (1996-2009). (c) and (d) compares the latitude-time cross-section of climatological mean daily precipitation and 850-hPa zonal wind over the Indian region (70-80°E) and (e) and (f) over the East Asian region (120-130°E).

from the same analysis period of June-August for 14 years (Fig. 9). In the comparison of the amplitude of the diurnal rainfall (Figs. 9a, b), the simulation is consistent with the observation in a sense that the signal is clear in the continents and adjacent oceans. Except the region of Tibetan Plateau, where the signal of diurnal cycle is much stronger than the observed, GloSea4 tends to underestimate the amplitude (for example, the tropical Maritime Continent, Southeast of the United States, and the northeast of Brazil). More striking difference is in the phase (Figs. 9c, d), where the seasonal prediction system exhibits too early convection both over the continents and oceans. Not shown here, most of land areas between 20°S-40°N shows the peak time in the late afternoon to early evening (1700-1900 LST) in the observations, whereas

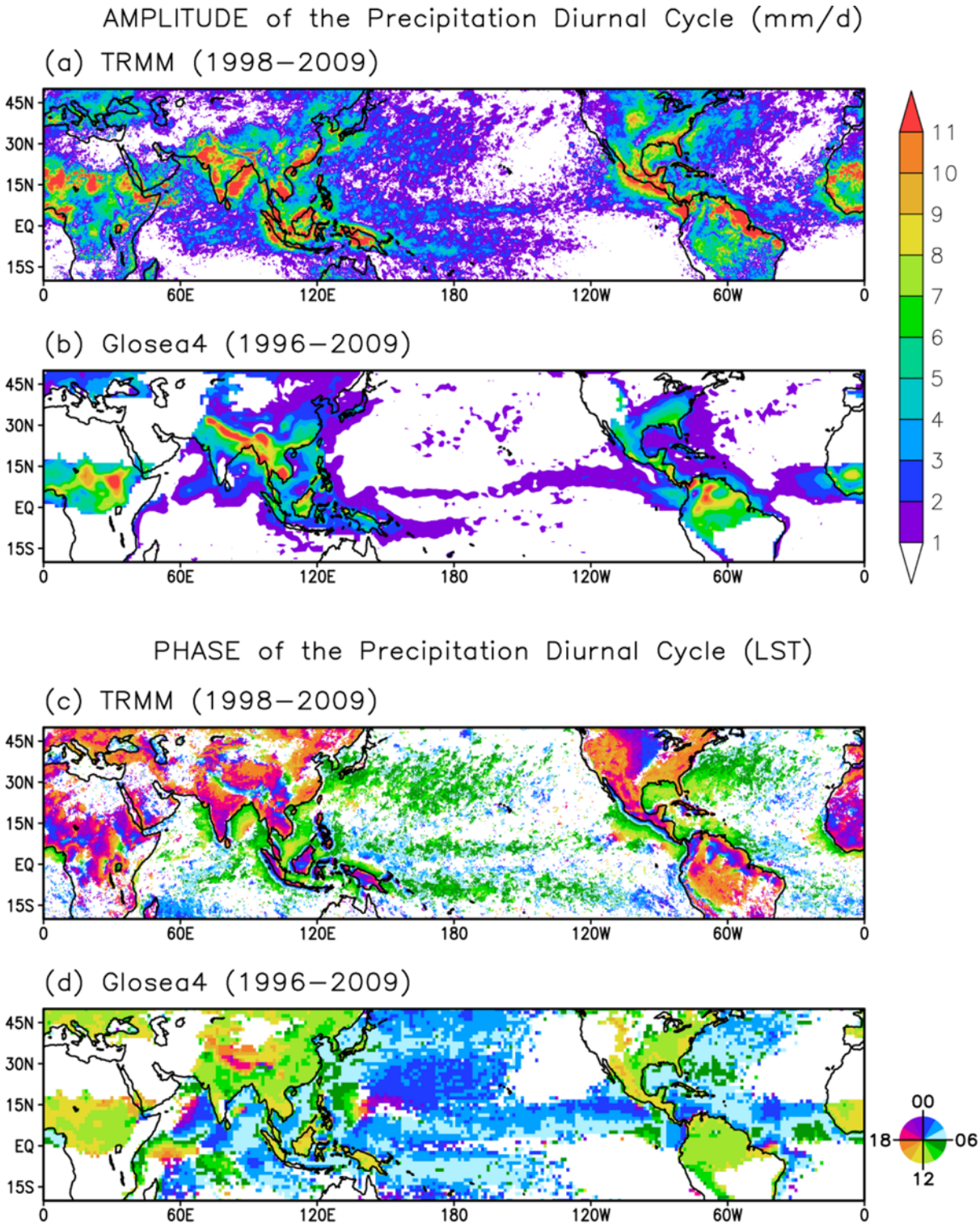

Fig. 9. Amplitude (a, b) and phase (c, d) of the diurnal cycle of summer (June-August) precipitation. (a) and (c) are from TRMM observations for 12 years (1998-2009) and (b) and (d) from GloSea4 for 14 years (1996-2009). Regions with statistically significant at 95% level are shaded.

that in the GloSea4 simulation shows a peak around the noon time (1100-1300 LST). It is noted that this bias is also evident in the winter. Regardless of season, it is a substantial bias in the rainfall, roughly 6 hours earlier than the observed.

The simulation of the diurnal cycle of precipitation in warm season is frequently examined for a purpose of diagnosing the

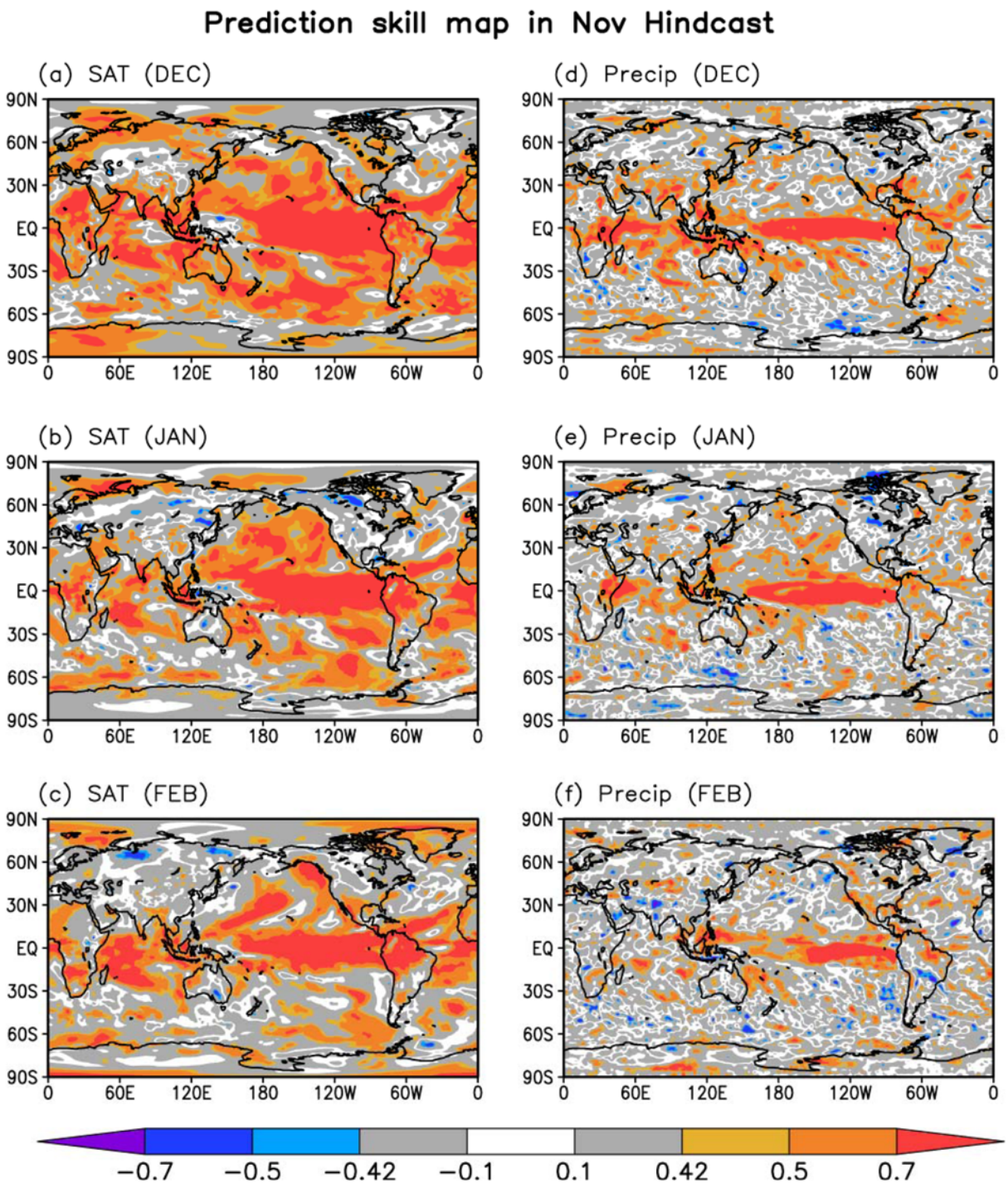

Fig. 10. Changes in the prediction skill as measured by anomaly correlation coefficient at the increase of forecast lead time. Left column is for the surface air temperature, and the right is for precipitation. Runs initialized in November are averaged for boreal wintertime prediction. Grey shading shows the area statistically insignificant at the 90% level.

physical parameterization of the model, particularly the moist convection process. Even in modern atmospheric reanalyses where the global prediction model is tightly constrained by observed atmospheric data the phase bias in the diurnal cycle is prominent. Too early development of rainfall and hardly precipitating in the nighttime over land is a typical problem pointed out by previous studies (e.g., Lee *et al*., 2007). A wrong phase of the precipitation diurnal cycle might introduce the erroneous diurnal variation of cloud and radiation, possibly distorting feedback in the land surface and degrading the quality of seasonal prediction, though the possibility has not been addressed quantitatively in the literature.

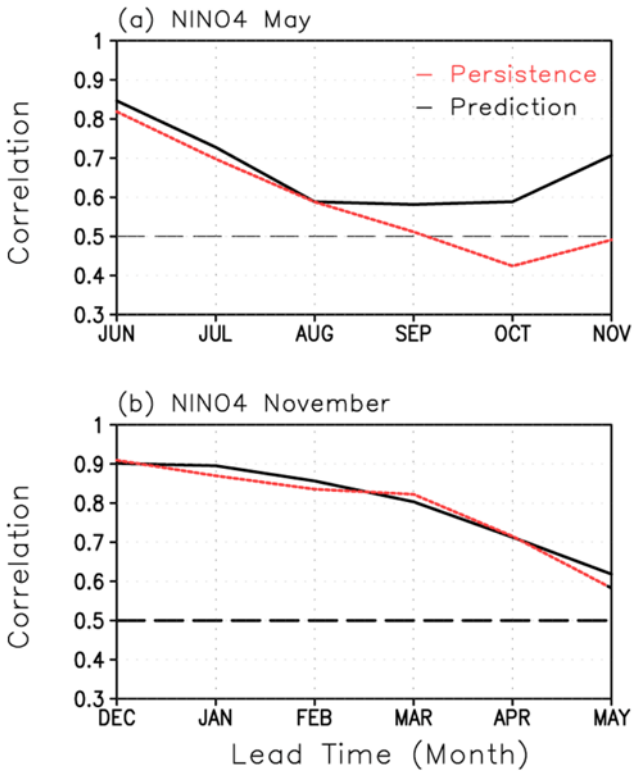

Fig. 11. Anomaly correlation skill of SST prediction by GloSea4 (black) in the Nino4 region. (a) shows the case of May initialized runs, and (b) the case of November initialized runs. The prediction skill is compared with the persistence curve (red).

4. Prediction skill of climate variability

a. Overall prediction skill

The overall prediction skill of surface air temperature and precipitation by GloSea4 ensemble hindcasts is evaluated in this section. The prediction skill is measured in this study by the anomaly correlation coefficient (ACC) between the observed and the predicted by GloSea4, which is one of the common choices used in the previous studies (e.g., Wang *et al*., 2009). We present the case for winter prediction as a typical case when the ENSO signal is stronger than other seasons. Note that the qualitative features of the prediction skill such as the geographical distribution of high skill and its changes with respect to the forecast lead time are more or less similar. Figure 10 shows the ACC map of surface air temperature and precipitation from the hindcasts initialized in November. We only show skills for the first three months after the month of initialization, because the prediction skill quickly goes away after the first three months particularly in the midlatitudes. In general, the model prediction for surface air temperature shows a better performance in oceanic area than that over the continental region. Gaining of the prediction skill is in fact mostly driven by high persistent nature of the sea surface temperature, rather than the actual skill (see Fig. 11). ACC tends to drop below the value of 0.42 (the 10% significance level) after the month 2 (for example, January in this plot) in the most of the continental region. This skill of seasonal prediction by GloSea4 is not much different from those performed by other seasonal prediction systems (e.g., Wang *et al*., 2009), indicating a current level of skill by global coupled model predictions. Note that the prediction skill is also poor over the maritime continent. This seems to be related with the relatively poor SST prediction skill in the western Pacific (e.g., the Nino 4 region), which will be discussed more in detail in Section 4b.

Figures 10d-f show the ACC map for precipitation. Overall, except the region of equatorial ocean where rainfall is strongly forced by ENSO, the prediction skill is quite poor in most of the area even in the month 1, suggesting that the prediction of precipitation is more challenging than that of surface air temperature. One can obtain a very similar feature in the boreal summer too (not shown), when a more accurate prediction is desired in monsoon areas including East Asia. Multi-model ensemble prediction shows similar skill for precipitation (Wang *et al*., 2009).

b. ENSO

The study of A11 showed that GloSea4 exhibits a decent skill of ENSO prediction as measured by ACC of the Nino 3.4 index (SST anomaly averaged over 5°S-5°N, 120°-170°W). ACC remains as high as 0.85 up to the 5 month lead time in the August forecast, for example. The results also showed the so-called "spring predictability barrier", where the prediction skill drops sharply as the forecast extends into the boreal spring time. We basically expand the analysis into the other indices, Nino 3 (5°S-5°N, 150°W-90°W) and Nino 4 (5°S-5°N, 160°E-150°W).

In our investigation to all season hindcasts, ACCs of three forecasted Nino indices remain above 0.6 throughout the forecast lead time up to 6 months (not shown). This suggests that the ENSO prediction by GloSea4 in two seasons ahead should remain useful. ACC for Nino 3.4 at 6-month lead time is about 0.75, which is fairly high when compared to other seasonal forecast systems such as SNU and MPI model (Jin *et al*., 2008). Among the three indices investigated, the GloSea4 system predicts Nino 3.4 index slightly better, and the prediction skill of Nino 4 is slightly lower than others. This result is somewhat contrary to the result of Hendon *et al*. (2009) who showed a better skill of the Nino 4 index than that of Nino 3 at lead times beyond 3 months. In our study, the prediction skill of Nino4 is almost comparable to that of Nino 4. The correlation skill for Nino indices exhibited a large spread depending on years, and the resulting statistics may depend on the validation period.

Figure 11 compares the skill of Nino 4 index between the hindcasts initialized in May and those initialized in November. As indicated in A11, the prediction skill is quickly dropping in the May runs (Fig. 11a). At this time of the year, the SST persistency in the observation in this region is also low (as

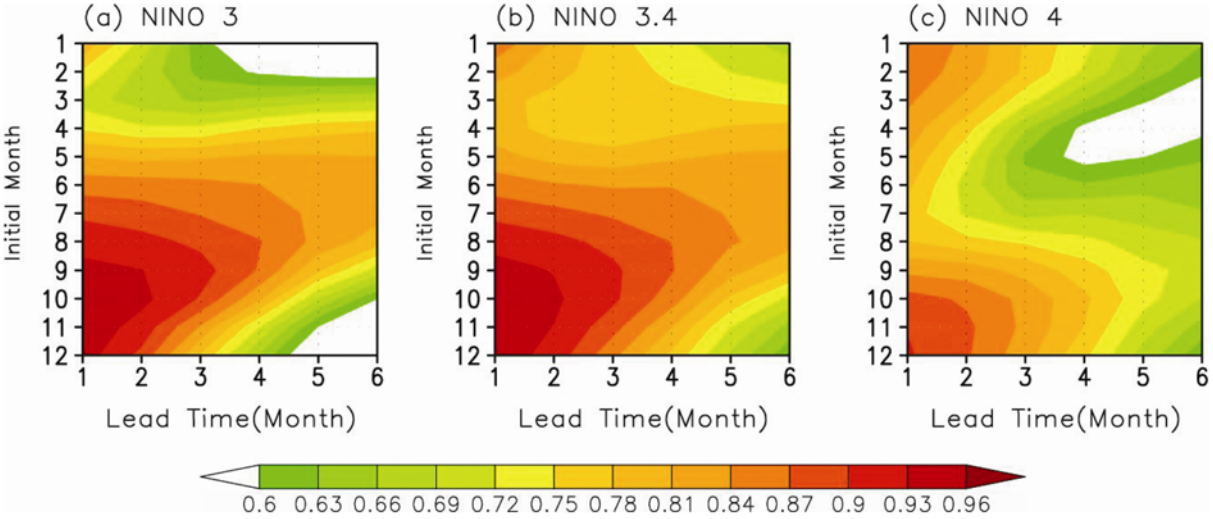

Fig. 12. Correlation coefficient between predicted and observed SST anomalies for (a) Nino 3, (b) Nino 3.4 and (c) Nino 4 regions indicated as a function of initial month and lead-time.

indicated in red line), which makes the forecast more difficult. GloSea4 demonstrates the skill over the persistence forecast, and ACC remains over 0.6 thought the forecast lead time. On the other hand, ACC remains higher apparently in the hindcasts initialized in November, but this is mostly due to the persistency of SST in the observations and the actual prediction skill is not significantly different from the persistence forecast.

Figure 12 shows the ACCs calculated as a function of forecast lead time (abscissa) and the initial month (ordinate). The plots indicate the prediction skill change at the increase of the forecast lead time (horizontal direction), as well as the forecast skill depending on the initialized month (vertical direction). It is shown that the GloSea4 system has a relatively higher prediction skill for the Nino indices when the system is initialized in September to January. The decrease of ACC and the increase of root mean squared error (not presented) as a function of lead time is most slow when the system is initialized in these months. On the other hand, when the system is initialized in March and April, ACC drops quickly, and RMSE increases rapidly, which again indicates problem of the spring predictability barrier. In general, the central tropics (Fig. 12b) show a better prediction skill than others. For Nino 3.4, correlation coefficient is higher than 0.69 even with initial months of March and April. Compared with ECMWF forecast system, ENSO prediction skill on all initial months is comparable with (slightly lower than) that presented in Kim *et al.* (2012). Also note that there is a difference in the timing of the prediction barrier across the regions. The prediction barrier is more pronounced and degrades the forecasts that initialized during the summer months over the Nino 4 region. SST anomaly over this region is known to have a considerable relation to the tropical storm activity (Emanuel, 2000), and this poor prediction skill of Nino4 index might have a potential to degrade the seasonal prediction skill of the tropical storm activity in summer season.

c. East Asia monsoon

Figure 13 shows the prediction skill of precipitation, surface temperature, and the 850-hPa zonal wind in the Asian monsoon domain during boreal summer. The geographical distribution of the prediction skill as measured by ACC is shown for each variable and for the first three months from June to August. These hindcasts were all started from May initial conditions. Even in the first month, the prediction of precipitation shows no good skill in most of the area except for the maritime continent and the western Pacific. Even these regions also experience a quick decay of the skill with the increase of the lead time. Throughout the season, most of the continent shows no statistically significant skill in the seasonal prediction of precipitation. At least for the first month prediction of surface air temperature, a broader region, particularly over the tropical oceans and western Pacific, shows a significant skill higher than 0.5, which is influenced by the persistence of the slowly-varying SST over the region. Coinciding with this signal, the 850-hPa zonal wind also exhibits a significant skill in the monsoon domain, which signal tends to last longer than other variables.

Being motivated by the case of 850-hPa wind prediction that shows an extended skill in summer, this study further examined the monsoon index prediction skill. This study tested two monsoon indices, one defined by Wang and Fan (1999) as the western Pacific East Asian monsoon index (WPEMI) and another defined by this study as the East Asian summer monsoon index (EASMI). WPEMI is defined as the 850-hpa wind speed anomaly difference between the average of 5-15°N and 100-130°E and that of 20-30°N and 110-140°E. EASMI uses the 850-hPa zonal wind only and it is defined as the wind anomaly difference between the average of 5-10°N and 130-

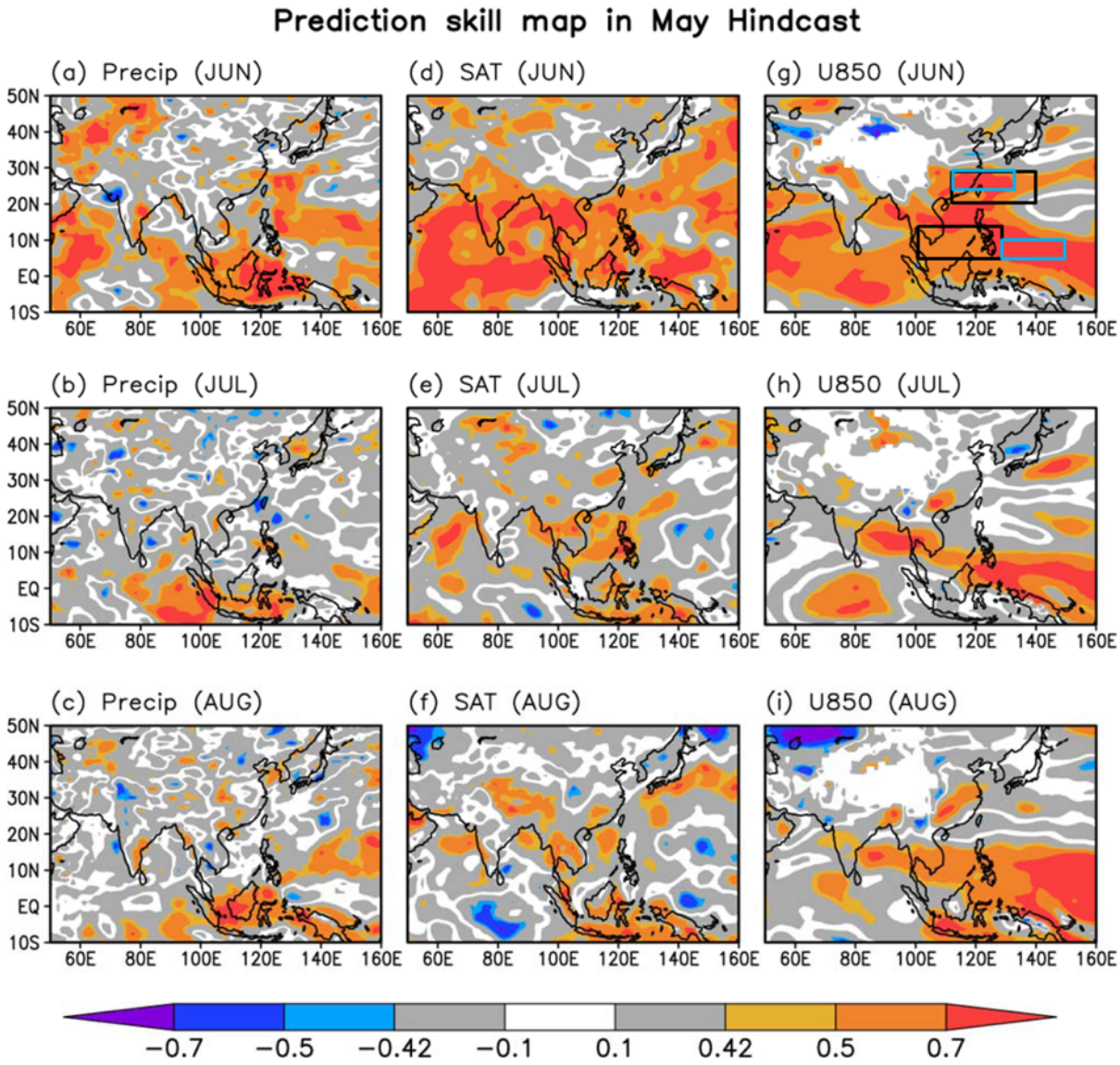

Fig. 13. Prediction skill of precipitation (left), surface air temperature (middle) and zonal wind at 850 hPa (right) between GloSea4 and observation (precipitation: GPCP, surface air temperature: MERRA and U850: ERA-interim). Grey shading shows the area statistically insignificant at the 90% level. Black boxes in (g) indicate the wind average region for the WPEMI monsoon index in Fig. 15, and blue boxes indicate the regions for the EASMI monsoon index.

150°E and that of 25-30°N and 110-130°E. The domain for EASMI was chosen so as to correspond to the high ACC region in Fig. 13g. Figures 14a, b show the observed WPEMI and EASMI and the predictions by GloSea4, respectively. The predictand is the JJA-mean monsoon indices from the May initialized ensemble hindcasts. As a result, the temporal correlation of WPEMI between the observation and the ensemble mean GloSea4 prediction is 0.42. This suggests that GloSea4 has a marginal skill for the WPEMI prediction. This correlation goes as below as 0.16 when the month-to-month values of WPEMI is predicted, implying that there seems much room to improve the prediction of sub-seasonal variability of East Asian monsoon. In the case of EASMI prediction, the GloSea4 prediction performs somewhat better than the prediction of WPEMI, with a higher temporal correlation of 0.67. This is a result somewhat expected by design, and it suggests that the dynamical prediction can be optimized further by utilizing past prediction skill from the hindcasts. Even if the dynamical model prediction suffers from systematic biases, an index-based prediction is still possible, such as the East Asia monsoon index prediction in this study. The index prediction is then translated using the observed relationship between the index and regional patterns of meteorological variables, based on any epoch composites or regression methods. This is the basis of the so-called “hybrid”-type seasonal prediction, which is one of the remedies to cure of the systematic bias of the dynamical prediction system.

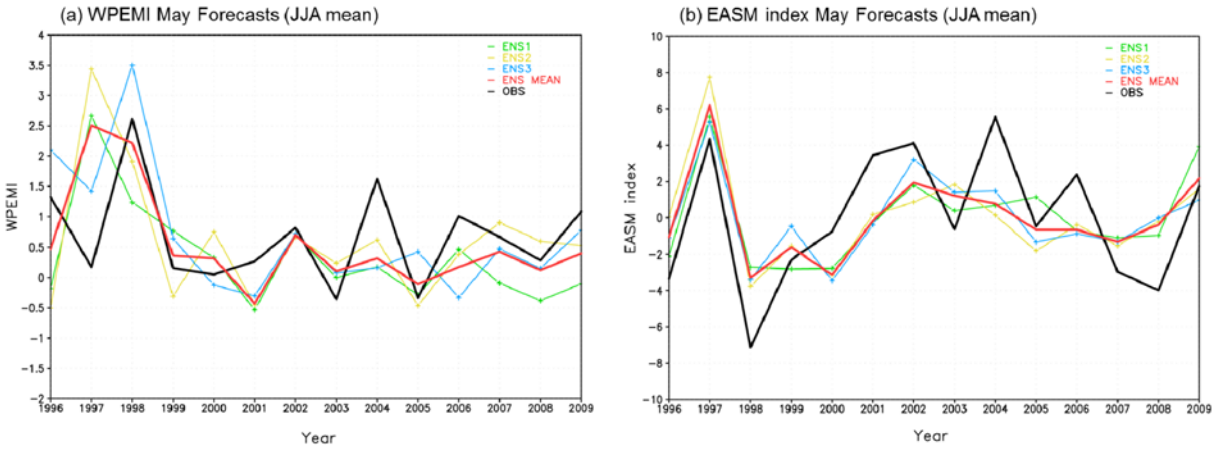

Fig. 14. The boreal summer (JJA) (a) WPEMI (Western Pacific East Asian Monsoon Index) and (b) EASM (East Asian Summer Monsoon) indices from observations (ERA-interim) and GloSea4. WPEMI (Wang and Fan, 1999) is defined as the wind speed anomaly at 850 hPa averaged over the region of 5-15°N and 100-130°E minus that over 20-30°N and 110-140°E. EASM is defined as the zonal wind anomaly at 850 hPa averaged over 5-10°N and 130-150°E minus that over 25-30°N and 110-130°E. The correlations between the observed and the GloSea4 predicted monsoon indices are 0.42 for WPEMI, and 0.67 for EASM.

d. Arctic Oscillation

The AO is known as an important seasonal climate variability of the East Asia especially during boreal winter. The definition of AO by Thompson and Wallace (1998) was used to examine the prediction skill by GloSea4, which is a leading EOF pattern of the monthly-mean sea level pressure (SLP) anomalies over north of 20°N during boreal winter (DJF). The AO index, as a predictand, is defined as the principal component (PC) of the 1st EOF (EOF1). Figures 15a, b are the SLP regression patterns associated with EOF1 from the MERRA reanalysis as the observation and the GloSea4 simulation. All 12 ensemble runs started from November (1-month lead hindcasts) were used for individual EOF analysis to SLP data, and Fig. 15b is the composite of those regression patterns. The SLP regression pattern from MERRA exhibits a dipole structure whose centers of action locate in the Arctic and the high latitudes, especially over the northeastern Pacific and Atlantic Ocean (Fig. 15a). The GloSea4 system has a robust pattern of AO with a clear dipole structure as in the observation. The variance of EOF1 is also comparable between the observed and the predicted, by explaining around 40% to the total wintertime variability. Figure 15c shows the PC time series from MERRA and GloSea4 hindcasts. Here we compare two hindcast results, one from the November starting runs (1-month lead hindcasts), and the other from the runs from September (3-months lead hindcast). To construct the forecast AO indices (time series), we used the EOF1 from MERRA for the time series projection in GloSea4 ensemble runs, although the results were not significantly different when the modeled EOF patterns were used respectively. Both AO PC time series display substantial year-to-year variation. Note that there was a strong negative phase of AO during 2009/2010 winter. The ACC between the observed AO index and the forecast AO indices are 0.52 for the 1-month lead hindcasts, and 0.26 for the 3-month lead hindcasts. This demonstrates that the GloSea4 has a significant prediction skill for the 1-month lead forecast, but the prediction is no good with 3 month lead time.

Figure 16 shows surface temperature and precipitation patterns regressed by AO index in MERRA and GloSea4. In the positive phase of AO, the observation (the MERRA reanalysis) shows a continental-scale pattern with a strong warming signal over the northern Siberia and eastern part of the North America, whereas anomalous cooling is in the Arctic, the northern part of the North America, northern Africa, Middle East and the Central Asia (Fig. 16a). The region of northern China, Korea, and Japan shows a warmer-than-normal state in the AO positive phase. The spatial structure of AO is reproduced remarkably well in GloSea4, although the magnitude is relatively weaker than the observed (Fig. 16b). Changes in the 200-hPa zonal wind according to the phase of AO (not shown) tend to explain much of this continental-scale response in the surface temperature anomaly, which feature is also reproduced well by GloSea4. Figures 16c, d compare the precipitation change associated with AO. Observed precipitation (Fig. 16c) shows higher (lower) precipitation in the Arctic and lower (higher) precipitation over mid- to high latitudes in the positive (negative) phase of AO. GloSea4 hindcasts (Fig. 16d) show very similar structure compared to observation. Note that the overall pattern of precipitation change simulated by GloSea4 is realistic as shown in this normalized pattern, but with much weaker amplitude in the non-normalized case. This suggests that, although the AO simulated by the GloSea4 system seems existing and physically consistent with the observed, its year-to-year variability is much weaker than the observed.

e. Madden-Julian Oscillation

The MJO is an eastward propagating, planetary-scale, sub-seasonal variability of tropical convection. Many studies have

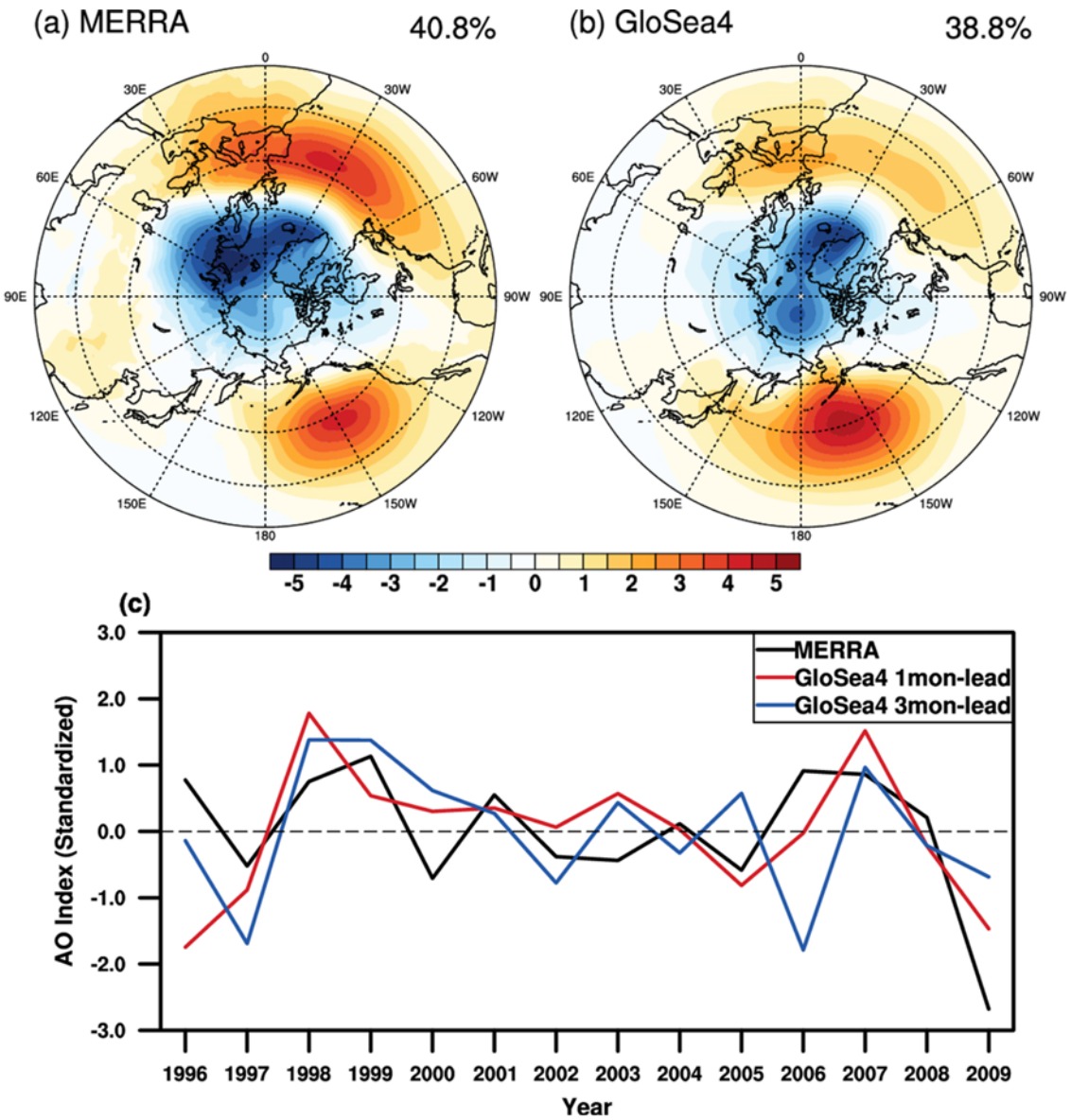

Fig. 15. DJF mean sea level pressure anomaly regressed onto leading PC for 1996-2009. (a) is MERRA, (b) is GloSea4 hindcast with 1-month lead time (unit is hPa). Percentages in right string indicate explained variance (averaged explained variance from each ensemble member) from the pattern. (c) is the DJF-mean normalized AO index from MERRA (black), and GloSea4 (red). The result from the GloSea4 hindcast with 3 month lead time is also presented (blue).

suggested the importance of MJO in affecting mid-latitude low-frequency variability in subseasonal time scale (e.g., Ferranti *et al*., 1990), including the recent study of Moon *et al*. (2011) who showed global teleconnection patterns associated with MJO and ENSO. For the evaluation of the MJO prediction skill in the GloSea4 hindcast data, the method of Wheeler and Hendon (2004) was adopted to obtain the MJO index. The Real-time Multivariate MJO (RMM) index of Wheeler and Hendon can be used for real-time MJO monitoring and forecasting, where the index is calculated using the combined EOF analysis with OLR and zonal wind at two levels of 850-hPa and 200-hPa. The other details of the analysis method are described in Wheeler and Hendon (2004). Note that for the forecast data, the previous 120-day mean is constructed by combining forecast data after the forecast starting date and observation data before the starting date. The RMM indices are obtained as the two leading principal components from the EOF analysis of the combined variable. In this analysis, we use spatial pattern obtained from Wheeler's website (http://cawcr.gov.au/staff/mwheeler/maproom/RMM), which is constructed using NOAA AVHRR OLR, and NCEP/NCAR zonal wind data. The patterns are used to construct forecasted RMM indices.

Figure 17 shows the prediction skill of the MJO by GloSea4 as assessed by two metrics: ACC and RMSE as a function of the forecast lead time. The calculation of the ACC and RMSE

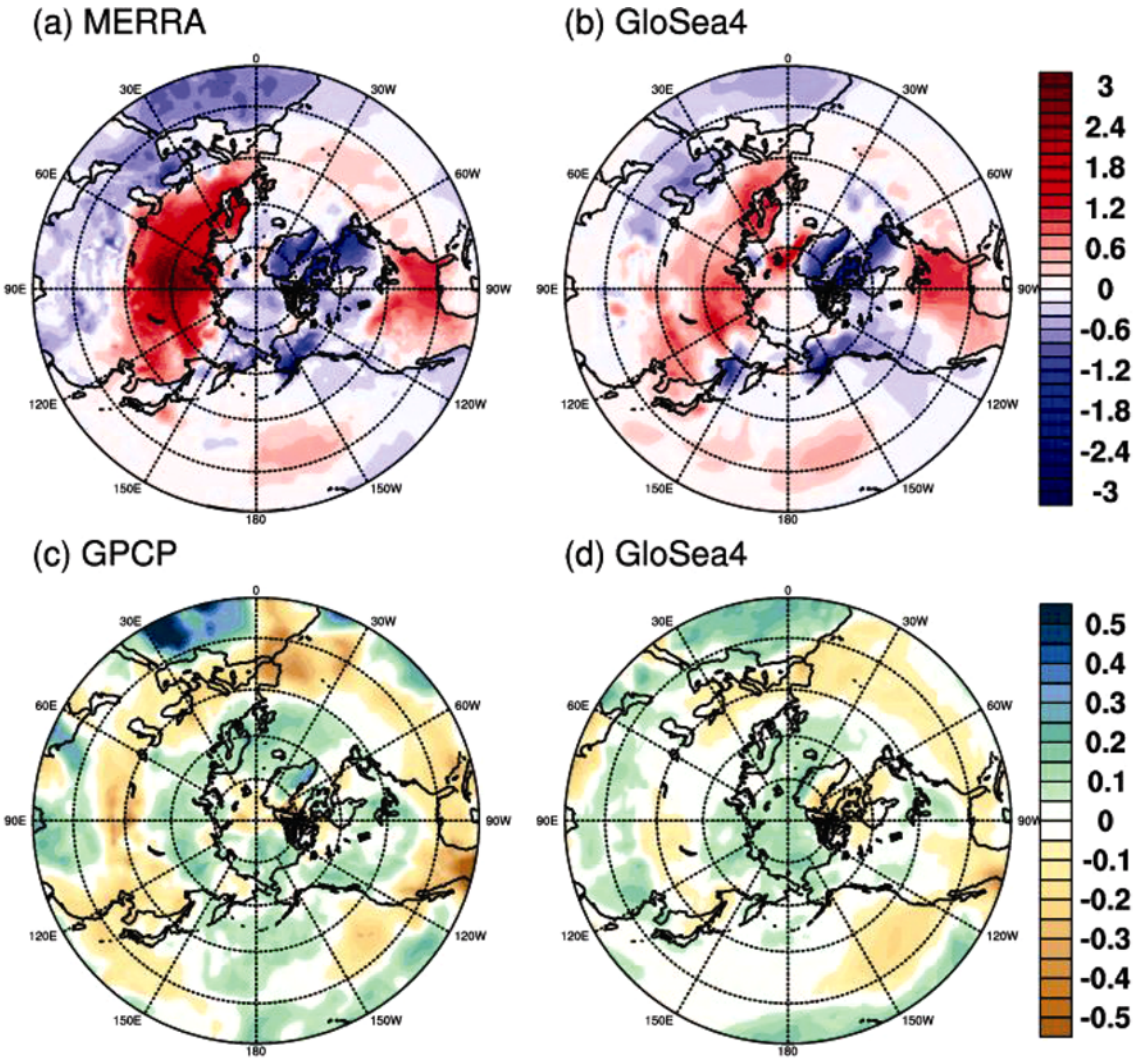

Fig. 16. Regression patterns of surface temperature (a-b, unit is K) and precipitation (c-d, unitless) by the AO index. Precipitation is normalized by time averaged precipitation of each grid. GPCP is used to validate of precipitation instead of MERRA.

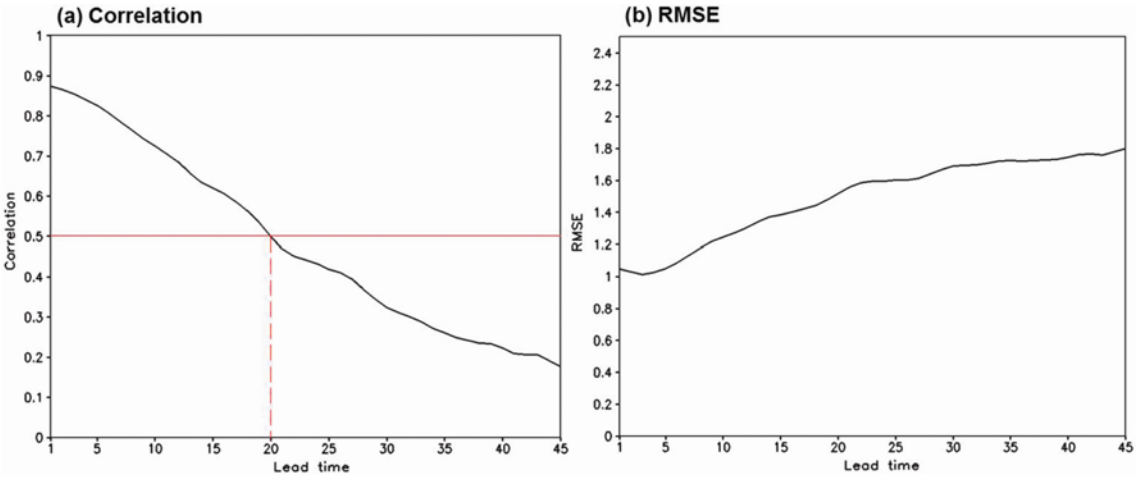

Fig. 17. (a) Prediction skill of MJO index by GloSea4 as measured by the correlation coefficient as a function of forecast lead time and the (b) root-mean-square error (RMSE). The red line indicates the 0.5 correlation.

followed that in Lin *et al*. (2008) and Gottschalck *et al*. (2010), and all available ensemble runs for 14 years were used to plot Fig. 17. The multi-year and multi-ensemble averaged ACC decreases at the increase of the forecast lead time, showing an useful forecast skill (ACC = 0.5) up to about 20 days. Note that this prediction skill is similar to what showed in A11 (their Fig. 17). Consistent with ACC, RMSE monotonically increases with the forecast lead time. Although not shown, there is substantial difference in the MJO prediction skill from year to year. For example, ACC drops to 0.5 in lead day 25 for year 1996, while it is about 13 day for year 2003. This suggests that the prediction skill of the MJO might depend on sea surface temperature condition in the tropics, although a further investigation is warranted.

4. Summary and conclusions

This study evaluated the basic performance of the seasonal prediction using a series of ensemble hindcast runs for 14 years (1996-2009) produced by the UK Met Office GloSea4 ensemble prediction system. GloSea4 is a state-of-the-art global prediction system for sub-seasonal to seasonal time scale, based on the fully-coupled HadGEM3 climate system model, with comprehensive initialization processes of atmosphere, land, ocean and sea ice. In this study, each ensemble hindcast run was initialized from the ERA-interim reanalysis for atmosphere and land surface. Ocean and sea ice states were also initialized by the ocean data assimilation.

To examine the fidelity of the system to reproduce and to forecast phenomena that are closely related to the East Asian climate, this study focused on two important aspects: quality of hindcast climatology and prediction skill of major climate variability. The first one includes the evaluation and the identification of the systematic biases in the global prediction model, judged from long-term averaged states with many ensembles. This diagnostic hopefully provides a useful guidance for the further improvement of the modeling system. The second one focuses on the prediction skill of ENSO, East Asian summer monsoon, AO, and the MJO, which are major drivers of weather and climate variability in East Asia.

The validation results have revealed that GloSea4 in general exhibits a remarkably good agreement with the observed climatology. A couple of systematic biases were identified, which are also commonly found in other climate models and prediction systems as well. The GloSea4 hindcasts show the energy imbalance at TOA and surface with a magnitude of less than 5 $W m^{-2}$. In addition, about 0.4 $W m^{-2}$ energy is lost internally in the atmosphere, which seems to be responsible for the overall cold bias in atmospheric temperature simulations. The cold tongue bias in SST in the eastern to central equatorial Pacific is pronounced, as like in other state-of-the-art climate models, and the more-than-observed SW reflection by cloud is speculated to be one of the causes and/or the feedbacks responsible for the bias.

Simulated climatology of precipitation captures the observed wet regions reasonably well, with a few model deficiencies including overall wet bias over the highly elevated region, and the split of ITCZs possible caused by the cold tongue SST bias. Zonal-mean precipitation is somewhat stronger than the observed, suggesting the modeled water cycle is stronger as well. GloSea4 exhibits a realistic East Asian summer monsoon simulation. In particular, the model is able to capture the seasonal, northward migration and retreat of rain belt associated with Changma, suggesting a potential to predict the Asian summer monsoon with a high fidelity.

The study also identified problems in the physical parameterizations, particularly in the deep convection scheme, requiring a further improvement of the global climate model in the seasonal prediction system. The analysis of the simulated hourly precipitation showed the substantial phase bias, where the modeled precipitation comes mostly in daytime and too earlier than the observation.

The overall seasonal prediction for surface air temperature and precipitation by GloSea4 shows a better performance in tropical ocean where SST has a high persistency. Prediction skill quickly drops at the increase of the forecast lead time in most of the continental region, with ACC below 0.42 in two months from the initialization. This poor performance in seasonal prediction is also common in other state-of-the-art prediction systems (Wang *et al*., 2009), defining overall status of the dynamical seasonal prediction.

Despite the system shows poor skill in predicting seasonal anomaly patterns in space, it nevertheless performs a statistically significant skill in the prediction of indices, such as ENSO, East Asian summer monsoon, AO, and MJO. This demonstrates that the prediction system has a capability of resolving observed major climate variability and predicting its temporal variation, in spite of the systematic biases in their forecast field in space. The GloSea4 system reasonably predicts ENSO indices at least 0.6 anomaly correlation coefficient with 6-month lead time. However, the spring barrier (i.e., the quick decrease of the ENSO prediction skill during boreal spring) is also pronounced as like in other systems such as CFS and ECMWF (Jin and Kinter, 2009; Kim *et al*., 2012).

The ACC skill score for the prediction of summer monsoon is also useful with a significant correlation of 0.67, providing a possible opportunity of the hybrid prediction that combines statistical and dynamical seasonal prediction. The wintertime AO index prediction based on the EOF analysis also demonstrates a useful skill. The ACC between the observed AO index and the forecast AO indices are 0.52 for the 1-month lead hindcasts, but the skill disappears in the 3 month lead forecast. This is somewhat contrary to the findings in A11, where the NAO prediction skill using the same GloSea4 system shows only a low correlation of 0.2. It is speculated that the AO prediction in this study can be substantially benefitted by the sea ice initialization in the Arctic Ocean, which was not implemented in A11 (see Section 2). The forecast skill of tropical MJO lasts up to about 20 days with ACC higher than 0.5. This is comparable to other prediction

systems (Kim *et al*., 2012).

Acknowledgments. The authors are grateful to the valuable comments from three anonymous reviewers. This work was supported by the Korea Meteorological Administration Research and Development Program under Grant APCC 2013-3141.

Edited by: Song-You Hong, Kim and Yeh

REFERENCES

Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-Present). *J. Hydrometeor.*, **4**, 1147-1167.

Arribas, A., and Coauthors, 2011: The GloSea4 ensemble prediction system for seasonal forecasting. *Mon. Wea. Rev.*, **139**, 1891-1910.

Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo, 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quart. J. Roy. Meteor. Soc.*, **134**, 1337-1351. doi: 10.1002/qj.289.

Bosilovich M. G., 2008: NASA's modern era retrospective-analysis for research and applications: Integrating earth observations. [Available online at http://www.earthzine.org/2008/09/26/nasas-modern-era-retrospective-analysis/.]

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Byun, Y.-H., and S.-Y. Hong, 2007: Improvements in the subgrid-scale representation of moist convection in a cumulus parameterization scheme: The single-column test and its impact on seasonal prediction. *Mon. Wea. Rev.*, **135**, 2135-2154.

Davey, M., and Coauthors, 2002: STOIC: A study of coupled model climatology and variability in tropical ocean regions. *Clim. Dynam*., **18**, 403-420.

Davies, T., M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N. Wood, 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **131**, 1759-1782.

Dee, D. P., and Coauthors. 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**: 553-597.

Emanuel, K. A., 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686-688.

_____, 2000: A statistical analysis of tropical cyclone intensity. *Mon. Wea. Rev.*, **128**, 1139-1152.

Essery, R. L. H., M. J. Best, R. A. Betts, P. M. Cox, and C. M. Taylor, 2003: Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *J. Hydrometeor.*, **4**, 530-543.

Ferranti, L., T. N. Palmer, F. Molteni, and E. Klinker, 1990: Tropical-extratropical interaction associated with the 30-60 day oscillation and its impact on medium and extended range prediction. *J. Atmos. Sci.,* **47**, 2177-2199.

Graeme, L. S., and Coauthors. 2008: CloudSat mission: Performance and early science after the first year of operation. *J. Geophys. Res.*, **113**, D00A18, doi: 10.1029/2008JD009982.

Gottschalck, J., and Coauthors, 2010: A framework for assessing operational Madden-Julian Oscillation forecasts: A CLIVAR MJO working group project. *Bull. Amer. Meteor. Soc.*, **91**, 1247-1258.

Gregory, D., J.-J. Morcrette, C. Jakob, A. C. M. Beljaars, and T. Stockdale, 2000: Revision of convection, radiation and cloud schemes in the ECMWF integrated forecasting system. *Quart. J. Roy. Meteor. Soc*., **126**, 1685-1710. doi: 10.1002/qj.49712656607.

Lin, H., G. Brunet, and J. Derome, 2008: Forecast skill of the Madden-Julian Oscillation in two Canadian atmospheric models. *Mon. Wea. Rev.*, **136**, 4130-4149.

Ham, S., and S.-Y. Hong, 2013: Sensitivity of simulated intraseasonal oscillation to four convective parameterization schemes in a coupled climate model. *Asia-Pac. J. Atmos. Sci*., **49**, 483-496.

Ham, Y.-G., and M. M. Rienecker, 2012: Flow-dependent empirical singular vector with an ensemble Kalman filter data assimilation for El Niño prediction. *Clim. Dynam*., **39**, 1727-1738.

Hastings, D. A., and W. J. Emery, 1992: The advanced very high resolution radiometer (AVHRR) - A brief reference guide. *Photogramm. Eng. Remote Sens.*, **58**, 1183-1188.

Hendon, H. H., E. Lim, G. Wang, O. Alves, and D. Hudson, 2009: Prospects for predicting two flavors of El Niño. *Geophys. Res. Lett*., **36**, L19713, doi:10.1029/2009GL040100.

Hewitt, H. T., D. Copsey, I. D. Culverwell, C. M. Harris, R. S. R. Hill, A. B. Keen, A. J. McLaren, and E. C. Hunke, 2011: Design and implementation of the infrastructure of HadGEM3: the next generation Met Office climate modelling system. *Geosci. Model Dev.,* **4**, 223-253.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, **8**, 38-55.

Hunke, E. C., and W. H. Lipscomb, 2010: CICE: The Los Alamos sea ice model documentation and software user's manual, version 4.1. LA-CC-06-012, Los Alamos National Laboratory, 76 pp.

Hwang, Y.-T., and D. M. W. Frierson, 2013: Link between the double-intertropical convergence zone problem and cloud biases over the Southern Ocean. *Proc. Natl. Acad. Sci*., **110**, 4935-4940.

Jin, E. K., and Coauthors, 2008: Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Clim. Dynam*., **31**, 647-664.

_____, and J. L. Kinter, III, 2009: Characteristics of tropical Pacific SST predictability in coupled GCM forecasts using the NCEP CFS. *Clim. Dynam*., **32**, 675-691.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, **37**, 173-199.

Kang, H. S., K. O. Boo, and C. Cho, 2011: Introduction to the KMA-Met office joint seasonal forecasting system and evaluation of its hindcast ensemble simulations. *36th NOAA Annual Climate Diagonostics and Prediction Workshop*, Fort Worth, TX, 3-6. [Available online at http://www.nws.noaa.gov/ost/climate/STIP/36CDPW/36cdpw-hkang.pdf.]

Kim, H. M., P. J. Webster, and J. A. Curry, 2009: Impact of shifting patterns of Pacific Ocean warming on North Atlantic tropical cyclones. *Science*, **325**, 77-80.

_____, P. J. Webster and J. A. Curry, 2012: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Clim. Dynam*., **39**, 2957-2973.

Koster, R. D., and Coauthors, 2010: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, **37**, L02402, doi:10.1029/2009GL-041677.

Kug, J.-S., Y.-G. Ham, M. Kimoto, F.-F. Jin, and I.-S. Kang, 2010: New approach for optimal perturbation method in ensemble climate prediction with empirical singular vector. *Clim. Dynam*., **35**, 331-340, doi:10.1007/s00382-009-0664-y.

Kusunoki, S., M. Sugi, A. Kitoh, C. Kobayashi, K. Takano, 2001: Atmospheric seasonal predictability experiments by the JMA AGCM. *J. Meteor. Soc. Japan,* **79**, 1183-1206.

Latif, M., and Coauthors, 2001: ENSIP: The El Niño simulation intercomparison project. *Clim. Dynam*., **18**, 255-276.

Lee, M.-I., and Coauthors, 2007: Sensitivity to horizontal resolution in the AGCM simulations of warm season diurnal cycle of precipitation over the United States and northern Mexico. *J. Climate*, **20**, 1862-1881.

Li, J.-L. F., and Coauthors, 2012: An observationally based evaluation of cloud ice water in CMIP3 and CMIP5 GCMs and contemporary reanalyses using contemporary satellite data. *J. Geophys. Res.*, **117**, D16105, doi:10.1029/2012JD017640.

Lin, J.-L., 2007: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean-atmosphere feedback analysis. *J. Climate*, **20**, 4497-4525.

Madec, G., 2008: NEMO ocean engine, Note du Pole de modélisation. Institut Pierre-Simon Laplace (IPSL), Tech. Note 27, 219 pp.

Maidens, A., A. A. Scaife, A. Arribas, J. Knight, C. MacLachlan, D. Peterson, and M. Gordon, 2013: GloSea5: The new met office high resolution seasonal prediction system. EGU general assembly 2013, 7-12 April, 2013 in Vienna, Austria, id. EGU2013-7649.

Mechoso, C. R., and Coauthors, 1995: The seasonal cycle over the tropical Pacific in coupled ocean-atmosphere general circulation models. *Mon. Wea. Rev.*, **123**, 2825-2838.

Misra, V., L. Marx, M. Brunke, and X. Zeng, 2008: The equatorial Pacific cold tongue bias in a coupled climate model. *J. Climate*, **21**, 5852-5869.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119. doi: 10.1002/qj.49712252905.

Moon, J.-Y., B. Wang, and K.-J. Ha, 2011: ENSO regulation of MJO teleconnection. *Clim. Dynam*., **37**, 1133-1149.

Park, S., S.-Y. Hong, and Y.-H. Byun, 2010: Precipitation in boreal summer simulated by a GCM with two convective parameterization schemes: Implications of the intraseasonal oscillation for dynamic seasonal prediction. *J. Climate*, **23**, 2801-2816.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609-1625.

Saha, S., and Coauthors, 2006: The NCEP climate forecast system. *J. Climate*, **19**, 3483-3517. doi: http://dx.doi.org/10.1175/JCLI3812.1

_____, 2013: The NCEP climate forecast system version 2. *J. Climate*, doi: abs/10.1175/JCLI-D-12-00823.1.

Schemm, J. E., K. C. Mo, and L. Long. 2011: Prediction of tropical storm season activities with the NCEP T382 CFS CGCM. AGU Fall Meeting 2011, abstract #GC23D-02.

Schneider, E. K., 2002: Understanding differences between the equatorial Pacific as simulated by two coupled GCMs. *J. Climate*, **15**, 449-469.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079-3102.

Spencer, R. W., 1993: Global oceanic precipitation from the MSU during 1979-91 and comparisons to other climatologies. *J. Climate*, **6**, 1301-1326.

Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190-1206.

Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297-1300.

Torrence, C., and P. J. Webster, 1998: The annual cycle of persistence in the El Niño /Southern Oscillation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1985-2004.

Valcke, S., 2006: OASIS3 user guide (prism_2-5) CERFACS technical support, TR/CMGC/06/73, PRISM report No 3, Toulouse, France, 60 pp.

Wang, B., and Z. Fan, 1999: Choice of South Asian summer monsoon indices. *Bull. Amer. Meteor. Soc.*, **80**, 629-638.

_____, and Coauthors, 2009: Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980-2004). *Clim. Dynam*., **33**, 93-117.

Webster, P. J., 1995: The annual cycle and the predictability of the tropical coupled ocean-atmosphere system. *Meteorol. Atmos. Phys*., **56**, 33-55.

_____, and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Quart. J. Roy. Meteor. Soc.*, **118**, 877-925.

Wheeler M. C., and Hendon, H. H., 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917-1932.

Wielicki, Bruce A., Bruce R. Barkstrom, Edwin F. Harrison, Robert B. Lee, G. Louis Smith, John E. Cooper, 1996: Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment. *Bull. Amer. Meteor. Soc.*, **77**, 853-868.

Yu, J.-Y., and C. R. Mechoso, 1999: Links between annual variations of Peruvian stratocumulus clouds and of SST in the eastern equatorial Pacific. *J. Climate*, **12**, 3305-3318.

Zhang, G. J., and H. Wang, 2006: Toward mitigating the double ITCZ problem in NCAR CCSM3. *Geophys. Res. Lett*., **33**, L06709, doi: 10.1029/2005GL025229.