

**On the comparison between seasonal predictive skill of global circulation models:
coupled versus uncoupled**

Asmerom F. Beraki,^{a,c*} Willem A. Landman,^{b,c} and David DeWitt^d

^a South African Weather Service, Pretoria, South Africa Private Bag X097, Pretoria, 0001,
South Africa

^b Council for Scientific and Industrial Research, Natural Resources and the Environment,
Pretoria, South Africa

^c Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South
Africa

^d National Oceanic and Atmospheric Administration, National Weather Service, Silver
Spring, Maryland, USA

*Corresponding Author, Email: asmerom.beraki@weathersa.co.za

Abstract

The study compares one- and two-tiered forecasting systems as represented by the South African Weather Service (SAWS) Coupled Model (SCM) and its atmosphere-only version. In this comparative framework, the main difference between these Global Climate Models (GCMs) resides in the manner in which the sea-surface temperature (SST) is represented. The models are effectively kept similar in all other aspects. This strategy may allow the role of coupling on the predictive skill differences to be better distinguished. The result reveals that the GCMs differ widely in their performances and the issue of superiority of one model over the other is mostly dependent on the ability to *a priori* determine an optimal global SST field for forcing the Atmospheric General Circulation Model (AGCM). Notwithstanding, the AGCM's fidelity is reasonably reduced when the AGCM is constrained with persisting SST

anomalies to the extent to which the Coupled General Circulation Model (CGCM)'s superiority becomes noticeable. The result suggests that the boundary forcing coming from the optimal SST field plays a significant role in leveraging a reasonable equivalency in the predictive skill of the two GCM configurations.

Key words: model comparison, AGCM, CGCM, Seasonal forecast, multi-model SST, sea-air interaction

1. Introduction

The practice of contemporary seasonal climate prediction requires state-of-the-art Global Climate Models (GCMs). The predictive skill of seasonal predictions mainly arises from the slowly evolving components of the climate system which are found to significantly modulate the mean state weather conditions [*Charney and Shukla, 1981; Palmer and Anderson, 1994; Barnston et al., 1999*]. Most of the signature of these slowly evolving systems is believed to originate from the ocean and thus the interaction between the ocean and the atmosphere is of paramount importance in the context of seasonal forecasting [*Goddard et al., 2001*]. In fact, GCMs are classified into two distinct configurations, commonly referred to as one- and two-tiered forecasting systems. These configurations are based on the manner in which information flows between the ocean and the atmosphere. In the atmosphere-only configuration, the Atmospheric General Circulation Models (AGCMs) are forced with independently predicted or persisted SST (sea-surface temperature) anomalies [*Bengtsson et al., 1993; Graham et al., 2005; Kug et al., 2008*] with the assumption that the atmosphere responds to SST but does not in turn affect the ocean [*Copsey et al., 2006*]. On the other hand, in one-tiered forecasting systems, the way in which the ocean and atmosphere interact and evolve mimics processes found in nature [*Palmer and Anderson, 1994*]. Therefore, this

spontaneous two-way feedback mechanism provides Coupled (ocean-atmosphere) General Circulation Models (CGCMs) a distinctive advantage over AGCMs [*Graham et al., 2005*].

Historically, two-tiered forecasting systems were the first to appear on the scene as seasonal forecasting tools and are still practiced globally [*e.g. Kirtman et al., 1997; Graham; 2000; Tennant, 2003; Beraki et al., 2015*]. Despite the enormous cost implications and complexity, one-tiered forecasting systems appear to have gained preference over two-tiered forecasting systems over recent years and their use by operational centers is steadily growing [*e.g. Stockdale et al., 1998; Palmer et al., 2004; Graham et al., 2005; Saha et al., 2006; Molteni et al., 2007; Beraki et al., 2014*]. This proliferation of interest is presumably stimulated by the fast development of computational technology complemented by the fact that many intercomparison studies suggest the superiority of CGCMs to AGCMs [*e.g. Yu and Mechoso 1999; Fu et al. 2002; Graham et al. 2005; Kug et al., 2008; Landman et al., 2012; Chaudhari et al., 2013; Zhu and Shukla, 2013; Shukla and Zhu, 2014*], even though similar studies report that only marginal differences exist [*e.g. Boville and Hurrell, 1998; Jha and Kumar, 2009; Colfescu et al., 2013*]. Most of these numerical studies report the weakness of AGCMs in simulating the Asian monsoon during the austral winter, where air-sea interaction plays a significant role. In contrast, CGCMs are distinctively able to rectify the problem and to better represent air-sea coupling in the tropical Indian and western Pacific Oceans [*Fu et al., 2002; Kug et al., 2008*]. In addition, *Graham et al. [2005]* suggest that coupled models can provide substantial benefits for seasonal prediction not only in tropical regions, but also in the extratropics.

It is commonly believed that coupled climate models are placed at the highest hierarchy in the science of numerical modelling in terms of complexity [*Stockdale et al., 1998; Palmer et al., 2004*]. In theory, they are largely hypothesized to represent state-of-the-art of seasonal forecasting which inherently renders them convenient for operational seasonal climate

prediction purposes. Notwithstanding, it may also be important to consider whether two-tiered forecasting systems offer comparable levels of skills that are currently attainable by state-of-the-art coupled models [Troccoli *et al.*, 2008] on one hand, and the inhibiting factor of the computational requirement to operate such coupled systems on the other hand. The latter consideration may be of particular importance in developing countries with less advanced capabilities, and especially at operational centers within these countries tasked to produce real-time seasonal forecast output. Moreover, although, as noted earlier, both model configurations are used at a number of operational centers, their comparison on seasonal prediction in an operational environment is less explored. It is worth emphasizing that it may be beneficial to objectively assess the relative merit or limitations of these forecasting systems under a constrained resources scenario. The aim of this paper is, therefore, to undertake a performance comparison of one- and two-tiered forecasting systems where the AGCM is constrained by the lower boundary conditions derived from predicted SST anomalies of two CGCMs' forecasts in contrast to persisted or empirically predicted SSTs [*e.g.* Graham *et al.*, 2005] while the two systems share a great deal of similarities in other aspects. To achieve our goal, the South African Weather Service (SAWS) Coupled Model (SCM) [Beraki *et al.*, 2014] and its atmospheric version [Roeckner *et al.*, 1996; Beraki *et al.*, 2015] are used. These two forecasting systems are currently running operationally at the SAWS as part of a multi-model system [Landman and Beraki, 2012]. The notion is also tested under a perfect model framework [Colfescu *et al.* 2013] and persistence (an AGCM forced with persisted SST anomalies) [Graham *et al.*, 2005]. The former configuration eliminates differences due to model bias between the CGCM and AGCM and enables the isolation of the role of coupling. In this framework, the AGCM is forced with the CGCM (SCM) retroactive SST simulations.

The paper is organized as follows: In section 2, the experimental design is described. Results from composite and time series analyses are presented on section 3. In section 4, differences in the predictive skill of the CGCM and the AGCM are elucidated. A summary and conclusions are given in section 5.

2. Experimental Design

2.1 Description of GCMs

The study compares the SCM and its atmosphere-only version as mentioned above. The SCM is described in depth in *Beraki et al. [2014]* while we only briefly describe the model here.

The SCM couples the ECHAM4.5 AGCM [*Roeckner et al., 1996*] and the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 3 (MOM3) [*Pacanowski and Griffes, 1998*] using the Multiple Program and Multiple Data (MPMD) fully parallelized coupler paradigm [*Komori et al., 2008*]. Under this coupling framework, the atmosphere and ocean models are treated as standalone versions apart from the interface that handles the exchange of information between the models.

While the AGCM, as in the two-tiered experiment, uses T42 (triangular truncation at wave number 42) horizontal resolution and 19 unevenly spaced hybrid sigma layers, the OGCM (Ocean General Circulation Model) has a 0.58° uniform zonal resolution, with a variable meridional resolution of 0.5° between 10° S and 10° N, gradually increasing to 1.5° at 30° S and 30° N and fixed at 1.5° in the extratropics. In the vertical, the OGCM uses 25 layers with 17 layers in the upper levels between 7.5m and 450m.

2.2 Retroactive forecasts

In this comparison experiment, the fundamental difference between the GCMs, as noted earlier, arises from the manner in which the ocean and atmosphere interact with each other. The two systems remain nearly identical in other respects. In the CGCM experiment, the AGCM and OGCM exchange information per simulation day. The AGCM feeds the OGCM with heat, momentum, freshwater, and surface solar flux. The OGCM, in turn, feeds the AGCM SST information. The coupling strategy used in this configuration is an anomaly coupling on the AGCM side and full-field coupling on the OGCM side, meaning that the anomalous atmospheric fluxes are superimposed on the observed climatology as in *Ji et al. [1998]* and *DeWitt [2005]*. The ocean initial conditions are taken from the ODA (Ocean Data Assimilation) system produced at GFDL that employs an optimum interpolation scheme [*Derber and Rosati, 1989*]. However, it is worth mentioning that seasonal climate prediction skill may be dependent on the accuracy of ODA on which systems are initialized from [*Zhu et al., 2012*].

In the two-tiered experiment, however, the AGCM is constrained by the lower boundary conditions derived from the predicted SST of two CGCMs combined with equal weighting. The two CGCMs are the SCM [*Beraki et al., 2014*] and the NCEP CFS v2 (National Centers for Environmental Prediction, Climate Forecasting System Version 2) [*Saha et al., 2014*]. The benefit of the multi-model approach has been reported in many forecasting studies over recent years [*e.g. Krishnamurti et al., 2000; Palmer et al., 2004; Doblas-Reyes et al., 2005, Hagedorn et al., 2005; Kirtman et al., 2014*]. In addition, the SST uncertainty amplitude (lower and upper bounds) defined from this combination is also considered as separate forcing fields. To identify the uncertainty amplitude a slight perturbation was applied to the multi-model ensemble mean of the SST independently for all lead-months (lead-0 to lead-8) using empirical orthogonal function analysis (EOF) [*North, 1984*]. The first normalized EOF mode was retained to describe the uncertainty term assuming that the variance is best

explained by the most dominant EOF mode and subsequently subtracted from or added to the multi-model SST [Beraki *et al.*, 2015].

Although more emphasis is placed on the understanding of whether the AGCM is a viable option to the CGCM under a constrained computational resources scenario, it is worth emphasizing that the use of multi-model SST forcing into the AGCM's configuration deviates slightly from a perfect model framework [Colfescu *et al.*, 2013]. For scientific purpose, this approach may not cleanly eliminate differences between the CGCM and AGCM due to model bias or isolate the role of coupling. However, from the perspective of operational forecasts, the multi-model SST forcing may offer a better optimization option. For interest of quantifying the extent of the disparity in biases and skill differences, the AGCM is also constrained by the predicted SST derived from the CGCM only, as in Colfescu *et al.* [2013], and for convenience hereafter referred to as "AGCMc". Also included are model simulations performed with persisting observed SST anomalies taken from the Optimum Interpolation version 2 [Reynolds, *et al.*, 2002] as a lower boundary condition to the AGCM (hereafter referred to as "AGCMp"). The rationale of including the AGCMp experiment is to gain additional insight into the AGCM's predictive skill relative to the CGCM when the AGCM is independently configured from the influence of the CGCM(s), as this may delineate the lower limit of the skill of the AGCM. Furthermore, this forecast strategy was used in similar comparative studies [*e.g.* Boville and Hurrell 1998; Graham *et al.*, 2005].

The CGCM and the AGCM (also AGCMp and AGCMc) use the same atmospheric initialization strategy. The atmospheric initial conditions (ICs) are obtained from the NCEP/DOE (Department of Energy) Atmospheric Model Intercomparison Project (AMIP) II Reanalysis (R2) dataset [Kanamitsu *et al.*, 2002]. The NCEP/DOE atmospheric states are

transformed to the horizontal and vertical resolution (T42L19) of the ECHAM4.5 AGCM in a manner that maintains numerical and gravitational stability as explained in *Beraki et al.* [2014]. The only difference is that the lower layer atmospheric temperature over the ocean (atmosphere-ocean interface) is assimilated from the multi-model ensemble mean of the SST and from the GFDL-ODA ocean state for use in the AGCM and the CGCM respectively. This is done to minimize the imbalance between upper-ocean mass field and wind stress [*DeWitt 2005*]. The uncertainties that arise from the ICs are accounted for by taking 10 consecutive daily atmospheric states back from the forecast date in each month and year. The November hindcasts, for example, consider the ICs that extend over the 10-day period from October 26 to November 4 for 28 years starting from 1982 and ending in 2009. The combination of ocean state and atmospheric state gives rise 10 and 30 ensemble integrations of the CGCM and the AGCM respectively.

2.3 Observation data

The model surface and upper air data are compared against the respective observed data compiled from different sources. For the surface variables, rainfall and air temperatures are acquired from the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) [*Xie and Arkin, 1997*] and the Climatic Research Unit (CRU) [*Harris et al., 2014*] respectively. For pressure data analyses, the (NCEP/ DOE) [*Kanamitsu et al., 2002*] is used as a proxy for observation.

3. Climatological and temporal differences

The results presented in this section are taken from the coupled and uncoupled models' hindcast simulations for the 28 years from 1982 to 2009. In this study, the lead-time is defined from the starting month when the model is initialized. For example, hindcasts from

November ICs for NDJ (November-December-January) are referred to as zero month lead-time hindcasts, while hindcasts for December-January-February (DJF), with the same initial conditions, are made at a one month lead-time, and so forth.

First, we investigate the role of the oceanic evolution of sea-air interaction by zooming in on the equatorial Indo-Pacific (Asian monsoon) region. Since the region has become the subject of many similar numerical studies, as noted earlier, it may be used as a benchmark for comparative assessments here.

The composite analysis of rainfall during the austral summer (DJF) at a one-month lead-time for part of the global region that centers the equatorial Indian and Pacific Oceans is shown in Figure 1. The two models (CGCM and AGCM) capture the CMAP climatological distribution reasonably well. They also consistently manifest similar bias patterns, with the exception that the AGCM exhibits a greater dry bias over the western and central Indian Ocean and a wet bias over the western Pacific Ocean. The CGCM is more biased over the eastern Pacific region south of the Equator at about 120°W. The AGCMc also exhibits similar bias patterns to the AGCM although the absolute bias difference is slightly shallower than the AGCM (Figure 1h, 1i). Similar composite analysis for the austral winter (June-July-August; JJA; Figure 2) suggests that both the CGCM and AGCM are able, once again, to represent the observed spatial patterns of rainfall reasonably well. The analysis also reveals that the CGCM appears to overestimate daily rainfall over the eastern equatorial Pacific region while the AGCM is more biased over the eastern Indian Ocean and western Pacific Ocean adjacent to Australia at about 10°S. It is worth noting that the AGCM and the AGCMc hardly differ in terms of bias distributions as also shown in their area averaged root mean square error (rmse) differences (see Figures 1, 2) and all GCMs appear to be more biased during the austral winter than they are during the summer (Figure 1d, 1e).

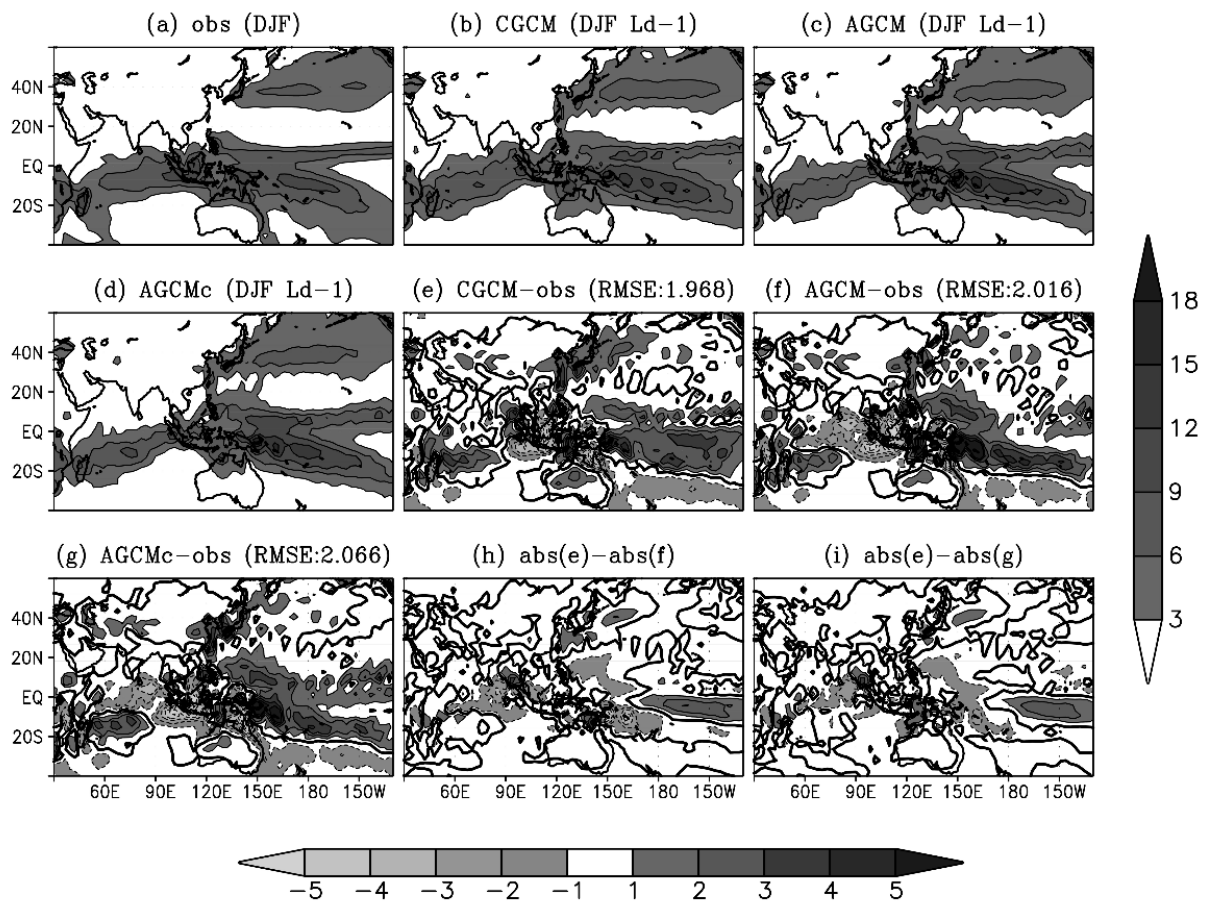


Figure 1. a Climatological representation of austral summer mean (DJF) precipitation (mm/day); (a) observation, (b) CGCM, (c) AGCM, (d) CGCM bias and (e) AGCM bias, (g) AGCMc (AGCM forced with the SCM predicted SST anomalies; see text) bias, (h) absolute value difference between CGCM bias and AGCM bias and (i) the same as (h) but for the AGCMc. The absolute bias differences enable to easily identify where exactly the GCMs differ, nonetheless the direction of the bias should be interpreted in conjunction to (e), (f) and (g). Also shown on the title of each bias plots is area averaged root mean square error (RMSE; e-g).

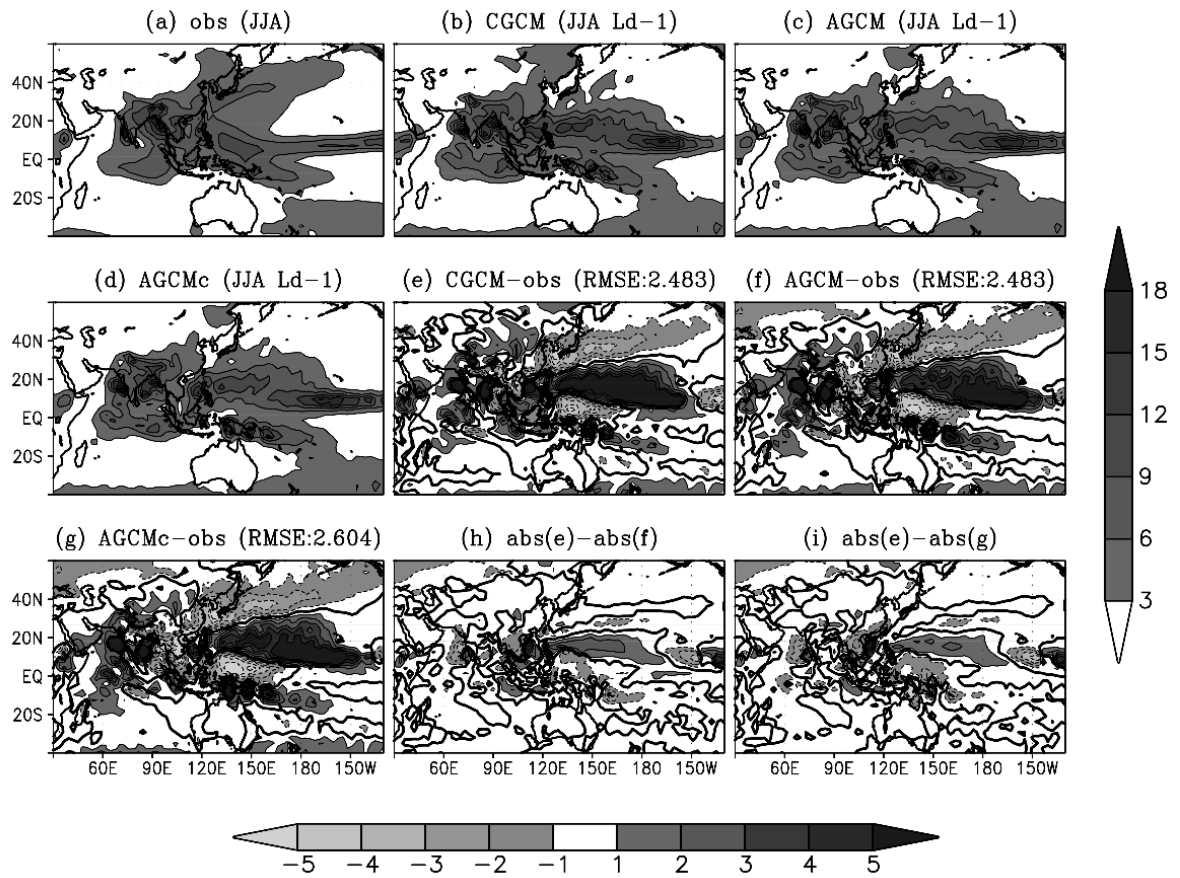


Figure 2. As Figure 1 but for austral winter.

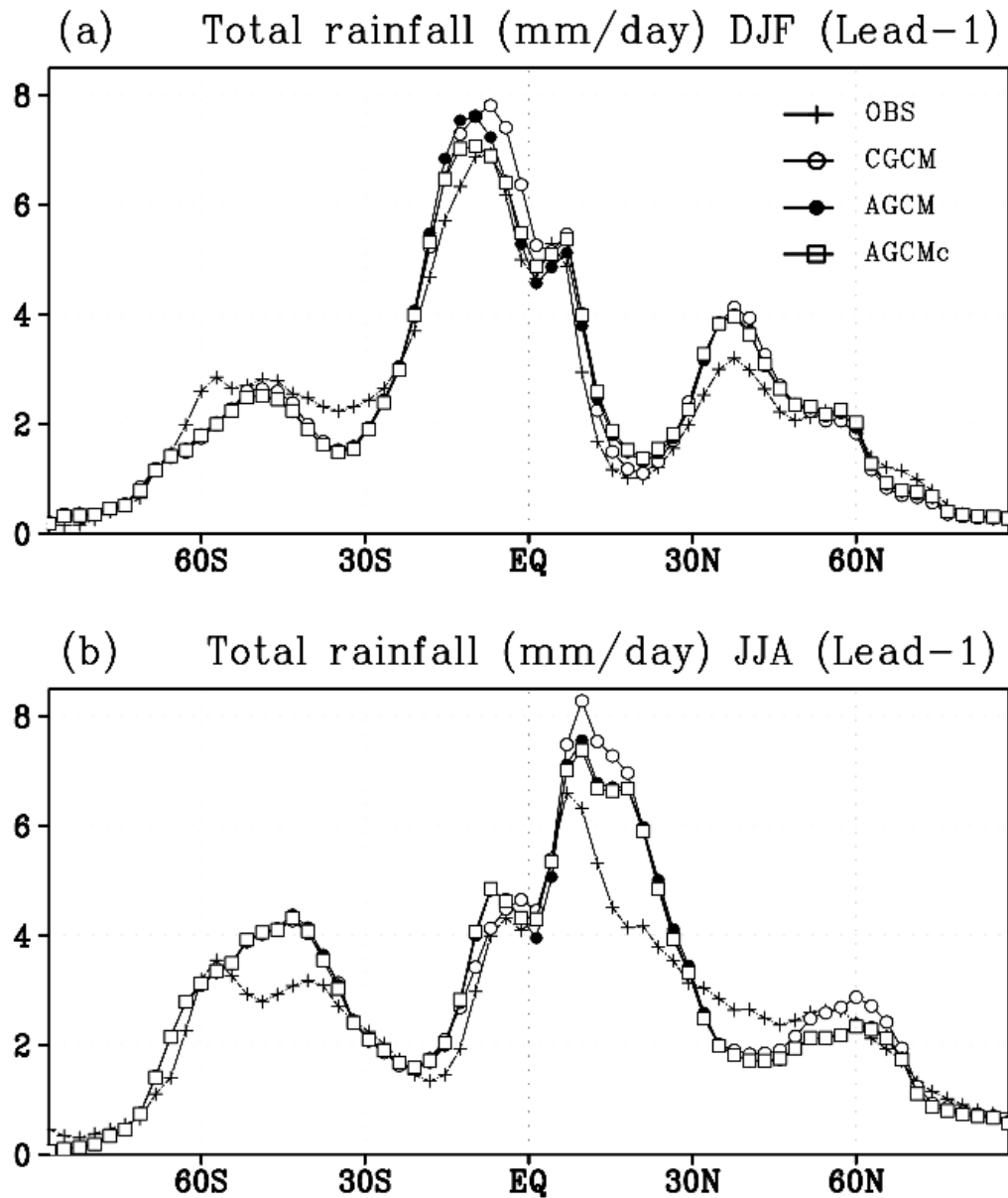


Figure 3. Zonally average total rainfall (mm/day) for DJF (a) and JJA (b) lead month 1 for the SCM, two different configurations of ECHAM 4.5 AGCM (as shown in the inset) and observation from CMAP [Xie and Arkin, 1996]. The temporal average is from 1982-2009 and the zonal extent is as in Figures 1, 2.

The zonally averaged (over the zonal extent of Figures 1, 2) DJF and JJA rainfall (mm/day) depicted in Figure 3 show that the CGCM, AGCM and AGCMc forecasts are in good agreement with the observed rainfall. In the tropics, the symmetry and position of the ITCZ (intertropical convergence zone) are well represented in all simulations. The mid-latitude storm tracks are also adequately represented despite all forecast strategies overestimate the DJF and JJA rainfall over the Northern Hemisphere (NH) and Southern Hemisphere (SH) respectively. However, the difference among the forecasting systems seems to be that the CGCM shows a slight tendency to exaggerate the tropical peak during DJF and JJA comparing to the two AGCM configurations.

The time evolution of rainfall biases across the equatorial Indo-Pacific region is further demonstrated in Figures 4-6 through the use of Hovmöller diagrams. The absolute bias differences computed from the GCMs biases and ENSO (El-Niño Southern Oscillation) information are also included in the plots to enhance objective interpretation and to better characterize the meridional and temporal bias differences. The ENSO phases are represented with the Oceanic Niño Index (ONI) [*L'Heureux et al., 2012*]. However, it is worth noting that care should be exercised when identifying wet or dry biases because the absolute value differences only emphasize whether the AGCM or the CGCM is more biased. During JJA season, the GCMs exhibit nearly indistinguishable differences in the Indian Ocean sector (Figure 4i, g) whereas the AGCM and the AGCMc consistently underrepresent the DJF CMAP estimates (Figure 4b - e). Notwithstanding, the GCMs biases are more pronounced over the equatorial Pacific Ocean than over the equatorial Indian Ocean. According to Figure 5, the AGCM exhibits wetter bias than the CGCM or the AGCMc during the austral summer over the western Pacific region. During the austral winter, in contrast, both AGCM prediction strategies (though more strengthened in the case of the AGCMc; Figure 5i, g) are relatively biased in comparison to the CGCM in the vicinity of the equator while the CGCM is

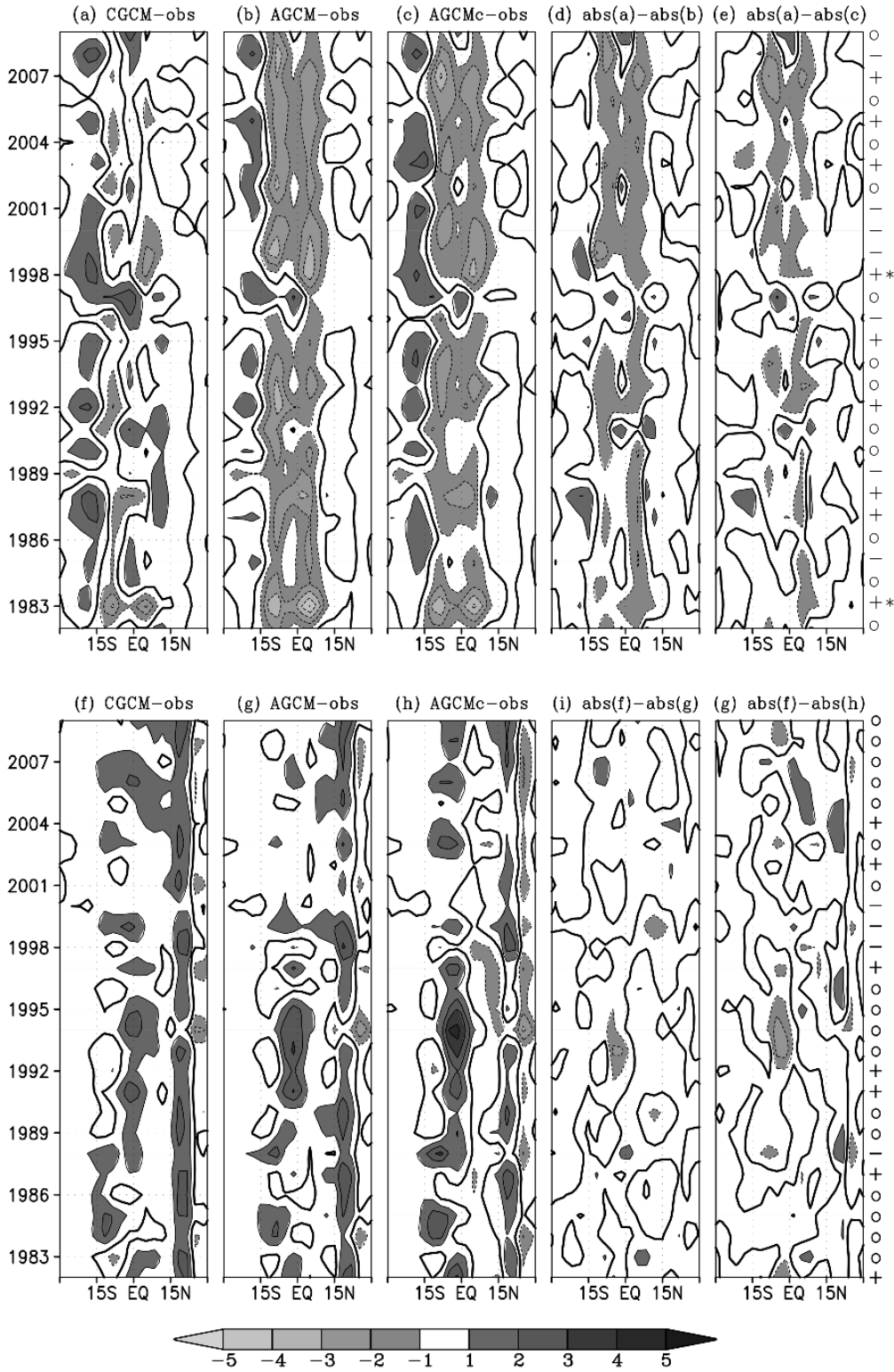


Figure 4. Hovmöller diagrams for the austral summer (top) and winter (bottom) rainfall at one month lead-time zonally averaged over the Equatorial Indian Ocean (50°E - 110°E). As shown in the title of each plot, rainfall biases are computed from the hindcast simulations of

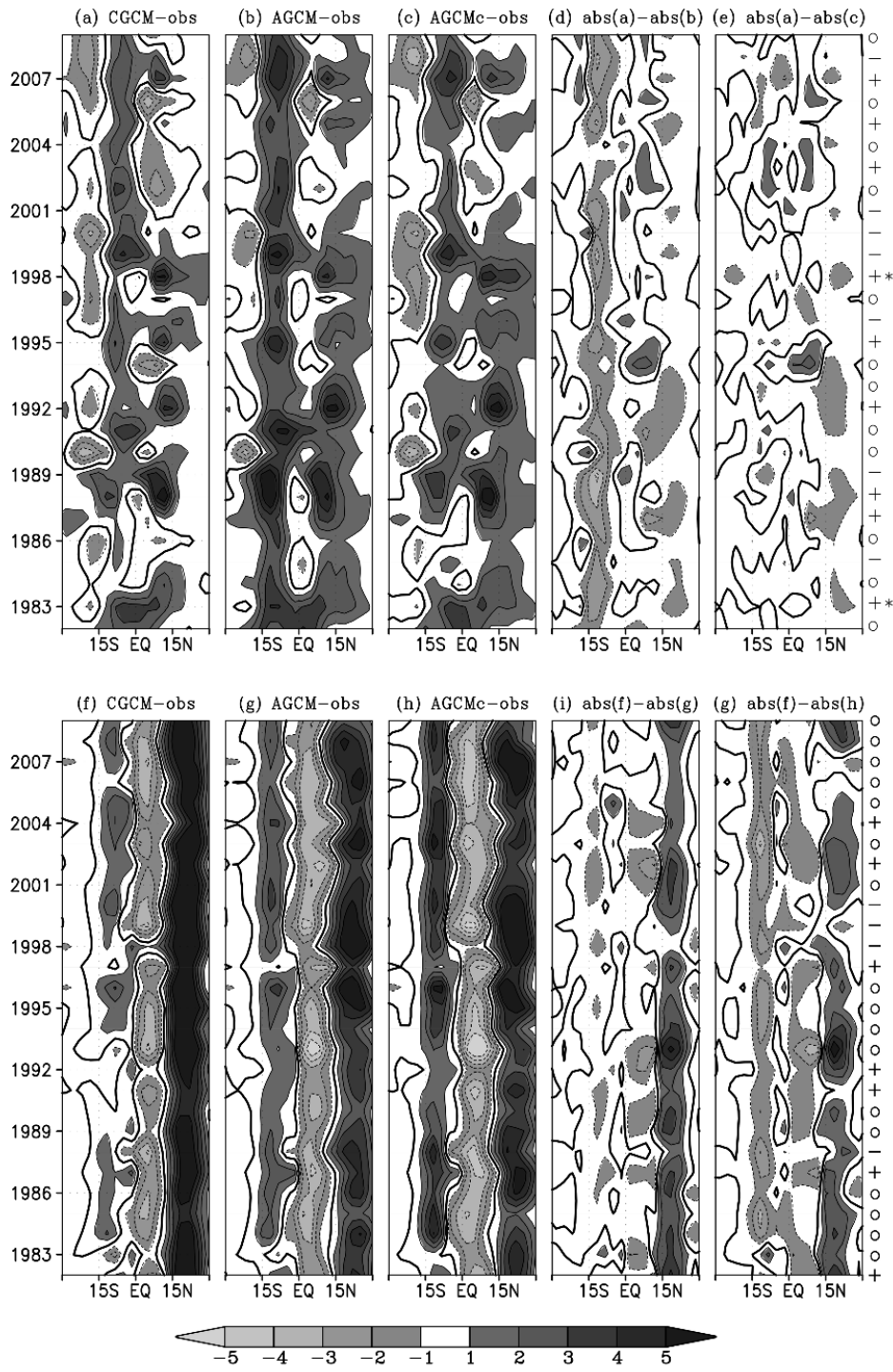


Figure 5. As in Figure 4 but for Western equatorial Pacific Sector (120°E - 170°E).

the GCMs against CMAP estimates and among the GCMs themselves. The absolute bias differences between the CGCM and AGCM (CGCMc) are also presented in (d), (e), (i) and (g) to highlight the meridional and temporal differences of the GCMs. The anomalous and neutral phases of ENSO, where - denotes La-Niña, + denotes El-Niño and o denotes neutral, are shown on the right side of plots (e) and (g). The two strongest El-Niño episodes are also indicated with *. The ENSO phases are based on the Oceanic Niño Index (ONI) [L'Heureux *et al.*, 2012] obtained from the National Oceanic and Atmospheric Administration (NOAA), Climate Prediction Centre (CPC; http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml).

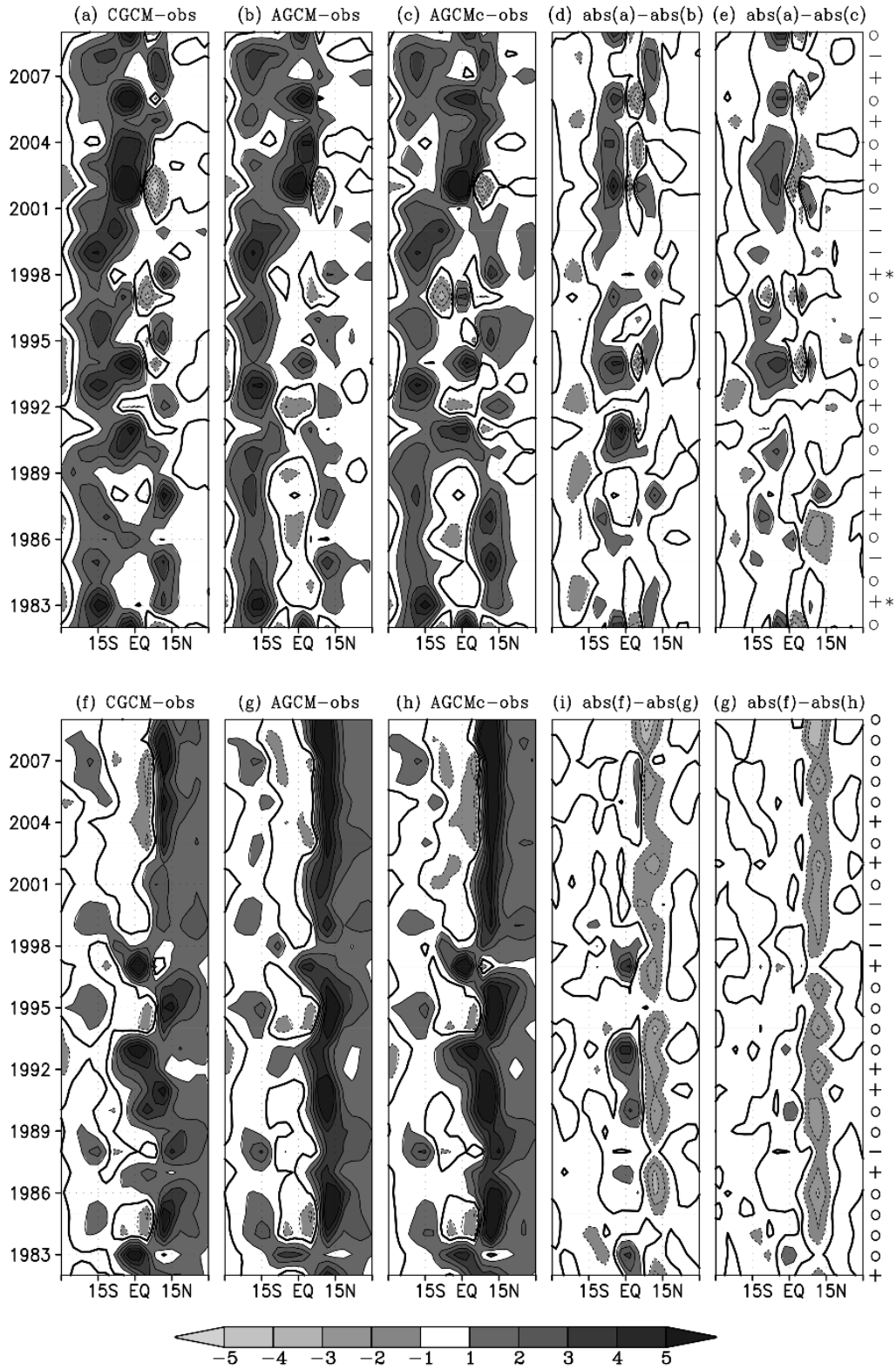


Figure 6. As in Figure 4 but for Eastern equatorial Pacific sector (170°E - 60°W).

noticeably biased over the southern hemisphere around 15°N. It is also noticeable that the CGCM overestimates the rainfall over the equatorial eastern Pacific during the austral summer (Figure 6d, e) while both AGCM configurations overestimate rainfall during the austral winter at about 15°N (Figure 6i, g), with the exception that the CGCM is more biased than the AGCM in 1983, 1989-1992 and 1997.

The biggest bias differences between the AGCM (AGCMc) and the CGCM over the equatorial Pacific regions mostly coincide with neutral ENSO. In the western Pacific during the austral winter season (Figure 5i, g), the largest biases of the CGCM are found (for example in 1989, 1990, 2001, 2009) during neutral ENSO years over areas surrounding 15°N; likewise, the AGCMc (largely reduced in the AGCM; Figure 5i) is noticeably biased (for example in 1984, 1985, 1988, 1993, 2003) over the area confined within the equator and 15°S. During DJF over the eastern equatorial Pacific, the CGCM biases are intensified in 1990, 1991, 2002 and 2009 (Figure 6d, e). Furthermore, the distribution of the model biases is more intense during the austral winter than during the summer, particularly for the western (Figure 4f-h) and eastern Pacific (Figure 5f-h). This result may suggest that the GCMs atmospheric response seems to depend on the accuracy of SST predictions. The dynamical ENSO predictive skill is minimized due to the NH spring barrier [*Saha et al., 2006; Beraki et al., 2014*]. The result further suggests that the CGCM's excessive rainfall in the tropics (Figure 3) may be attributed to the SST bias found within the time evolution of the air-seas coupling process which is slightly minimized in the AGCM (AGCMc) with the use of prescribed SST forcing. This may reinforce the conclusion that both GCMs are able to respond to the SST fluctuations equally and that the role of the evolution of sea-air interaction anticipated to favor the CGCM is not clearly established in the cases we have tested for the Asian monsoon regions, particularly during the austral winter despite that

previous similar studies reported in favor of the CGCMs [e.g. *Fu et al., 2002; Kug et al., 2008*].

However, the result presented above does not consider whether those differences between the two forecasting systems are statistically significant. To approach the problem indirectly, we perform a statistical significance test using the Wilcoxon–Mann–Whitney non-parametric approach without involving observations [*Graham et al., 2005; Wilks, 2006*]. Figures 7 and 8 show the spatial extent and temporal frequencies when the CGCM and the AGCM rainfall fields are found to be different in their probability density functions (PDFs) with a statistical significance at the 95% level for the austral summer and winter seasons at a one month lead-time respectively. The PDFs are represented with 10 and 30 ensemble members of the CGCM and the AGCM respectively. According to this result, more than 80% of the time the source of variation between the coupled and uncoupled models arises mostly from the equatorial region. During the austral summer, more pronounced differences are found over the equatorial eastern Pacific Ocean, Brazil, Atlantic Ocean and the south eastern Australian coast (Figure 7a). Likewise, these differences are noticeable in the Pacific Ocean and Atlantic Ocean off the coast of West Africa during austral winter (Figure 8a). Most of the peaks in the areal extent differences between the two models are consistently concentrated over the Pacific region during JJA although the differences are confined over the eastern Pacific region during DJF. It is also noticeable that the peaks are mostly found during neutral ENSO conditions. As noted earlier, the GCMs also differ mostly over the equatorial Pacific region in terms of biases (Figures 5 and 6). Similar analysis between the CGCM and the AGCMr (AGCMc), in which the GCMs use the same ensemble size, demonstrate that most of the source of differences similarly arise from the equatorial region. However, in both cases the differences in the areal extent over the tropical Pacific are reduced to some extent particularly

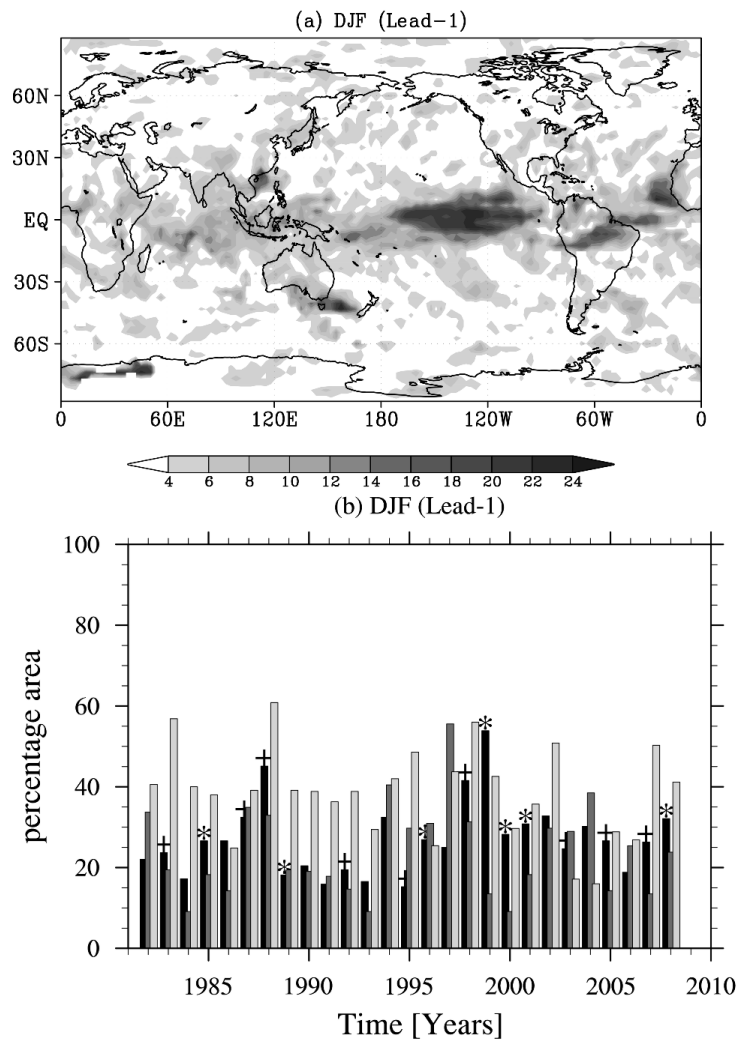


Figure 7. Number of years (out of 28, 1982–2009) when CGCM and AGCM ensemble distributions for DJF (lead-1) rainfall were found significantly different at the 95% level with a Wilcoxon–Mann–Whitney test performed at each grid point (a); time series of the percentage area where CGCM and AGCM ensemble distributions for austral summer (DJF Lead-1) rainfall are found significantly different at the 95% level (b). Black, dark grey, and light grey bars denote equatorial Indian Ocean, western equatorial Pacific Ocean and equatorial eastern Pacific Ocean respectively. Annotation represent anomalous phases of ENSO where * denotes La-Niña and + denotes El-Niño. The ENSO phases are based on the Oceanic Niño Index (ONI) [L’Heureux et al., 2012].

between the CGCM and the AGCMc during JJA seasons (not shown). The result attests the realism of the bias differences discussed so far.

4. Differences in the Seasonal Predictive skills

The GCMs are investigated for their relative performance for different seasons and lead-times. This comparative analysis is based on the 28 years hindcast of the AGCM and the CGCM which consist of 30 and 10 ensemble members respectively. Furthermore, results from the AGCMc and the AGCMp, which each has the same ensemble size as the CGCM, are also presented. Each ensemble set mimics a set of operational forecasts because the sets were created in a manner similar to how operational forecasts at SAWS, a WMO-recognized Global Producing Centre, are being conducted. This approach offers a better insight into the relative enhancement or degradation of forecast quality in an operational environment.

4.1 Comparison based on the ensemble mean

The impact of the oceanic evolution of sea-air interaction (only supported in the CGCM) and the use of prescribed multi-model SST forcing (only supported in the AGCM) on the predictive skill of seasonal forecasts is compared by first evaluating the ensemble mean of the GCMs. The anomaly of each model is computed about its own drifted climatology before the statistics are applied in order to remove biases from the model forecasts as a function of lead months. We first concentrate on atmospheric pressure fields as the signatures of most climate drivers (ocean-atmosphere coupling phenomena) including, inter alia, ENSO [*Neelin et al., 1998; Wallace et al., 1998*], Indian Ocean Dipole (IOD) [*Saji et al., 1999*] and Pacific South America (PSA) [*Mo and Ghil, 1987*], which are represented in the mean sea level pressure (mslp) or geopotential height (GH) fields. To facilitate the comparison, we use the mean square skill score (MSSS) [*Murphy, 1988*]. The MSSS is computed using the mean square

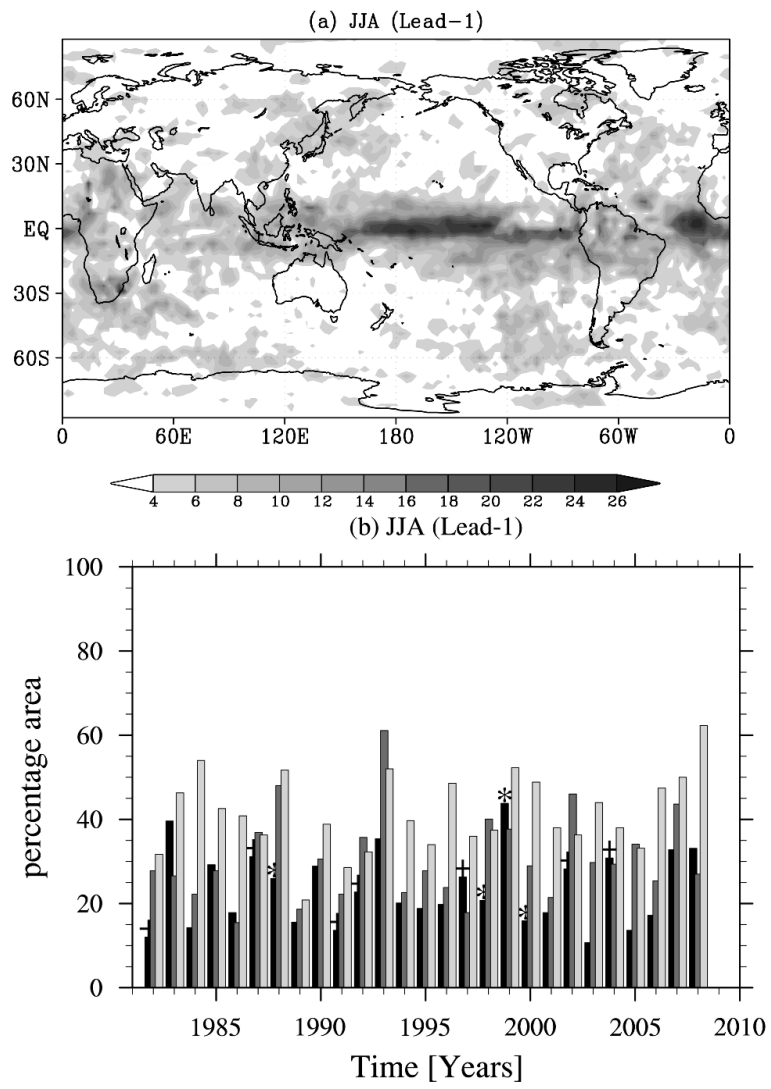


Figure 8. As in Figure 7 but for austral winter.

error (MSE). The MSE of each model is independently calculated first against the NCEP/DOE. In this comparison, the MSSS of the CGCM is identified by taking the AGCM as a reference forecast (i.e., $MSSS = 1 - \text{MSE}_{\text{CGCM}} / \text{MSE}_{\text{AGCM}}$). The skill score therefore represents gains or losses in the forecast skill relative to the reference forecast. In this context, the MSSS approaches +1 (approaches $-\infty$) if the CGCM (AGCM) perfectly outperforms its counterpart and the gradient between +1 and $-\infty$ measures the degree of superiority (inferiority) of one model over the other. The CGCM and AGCM predictive skill equates when the ratio of MSE of one model to the other approaches 1.

Figure 9 shows the extent to which the CGCM's predictive skill has improved (degraded) relative to the AGCM in predicting the mslp during DJF and JJA seasons with a 95% level of statistical significance. The significance level is identified using the bootstrap non-parametric procedure (sampling with replacement) [Wilks, 2006] where the analysis is repeated 1000 times. According to this result, the predictive skill of the CGCM during the austral summer appears to significantly strengthen across the equatorial Indian Ocean and in the vicinity of central and North America relative to the AGCM, but the benefit diminishes as a function of lead-time. The AGCM, however, consistently outperform the CGCM on the equatorial Pacific region across all lead-times considered. Results from AGCMc (maps not shown) demonstrate a relative skill degradation predicting mslp. The finding may support the notion that the use of multi-model SST forcing in the AGCM configuration played a significant role for the best performance of the AGCM. Beraki *et al.* [2015] also showed that the skill and accuracy of the Nino3.4, derived from the multi-model ensemble SST used to force the AGCM, intensified during the austral summer as the lead lead-time increases.

During the austral winter, the coupled model gains significant advantage over the atmospheric model particularly at one and three month lead-times on the equatorial region, with the exception of the eastern Pacific sector. The strength of the CGCM skill over the

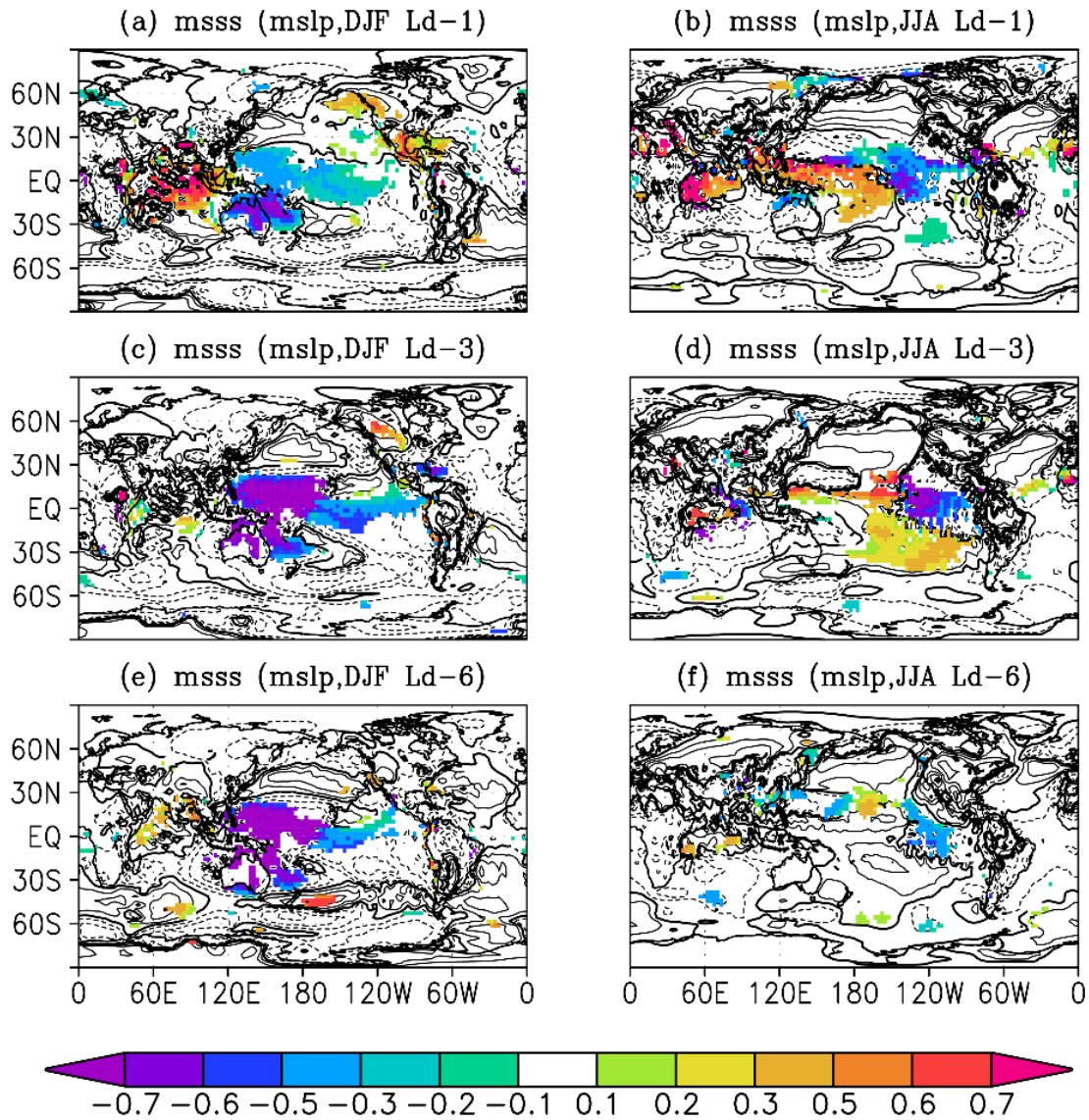


Figure 9. Skill improvement or degradation of the CGCM relative to the AGCM (reference) in predicting mslp (Pa) during the austral summer (December-January-February; DJF; left panel) and winter (June-July-August; JJA; right panel) for various months lead-time as shown in the title of each plot. The MSE of each model is first computed against the NCEP/DOE mslp that eventually returns the MSSS. The region with +ve (-ve) scores imply the superiority of the CGCM (AGCM) where those statistically significant at 95% level with reasonable differences are shaded. Also shown is contours with 0.2 interval. The significance test is performed with a bootstrap non-parametric procedure [Wilks, 2006].

equatorial Indian Ocean is also noticeable at short lead-times. However its skill is significantly reduced over the eastern part to the extent that the AGCM takes the lead as the lead-time increases (Figure 9d, 9f). In the prediction of the austral winter mslp over the Asian monsoon region, the CGCM outperforms the AGCM, although its superiority decays quickly as the lead-times increase.

In the 500hPa geopotential height (GH) comparative analysis, the AGCM's benefit, during the DJF season is well manifested over the eastern part of the equatorial Pacific region and the northern South America sub-continent with a tendency of deepening as the lead-time increases (Figure 10c, 10e), despite the CGCM performing better over the southern Africa sub-continent. At enhanced lead-times (Figure 10c, 10e) the dominance of the AGCM is also extended over equatorial Africa and the Indian Ocean. During the JJA season, the CGCM is found to outperform the AGCM over the Asian monsoon basin at a one month lead-time, although the benefit is changed in favor of the AGCM at extended lead-times. The superiority of one model over the other during the winter season is mostly indistinguishable as opposed to the austral summer for the 500hPa GH. Previous evaluation studies [Beraki *et al.*, 2014; 2015] conducted on these models using the same hindcasts show independently that reasonable skill of the GCMs in predicting pressure fields taken climatological forecast as a reference is found mostly over the equatorial region and the predictive skill presented here should therefore be viewed in relative terms.

The finding is consistent with what has been discussed so far with regard to the prevalence of noticeable differences over the equatorial (notably Pacific) region in spite of the narrowing tendency in (bias and skill) differences under the perfect model framework. The other point worth mentioning is that the evolution of sea-air interaction expected to favor the CGCM is barely supported particularly at longer lead-time. This result may suggest that

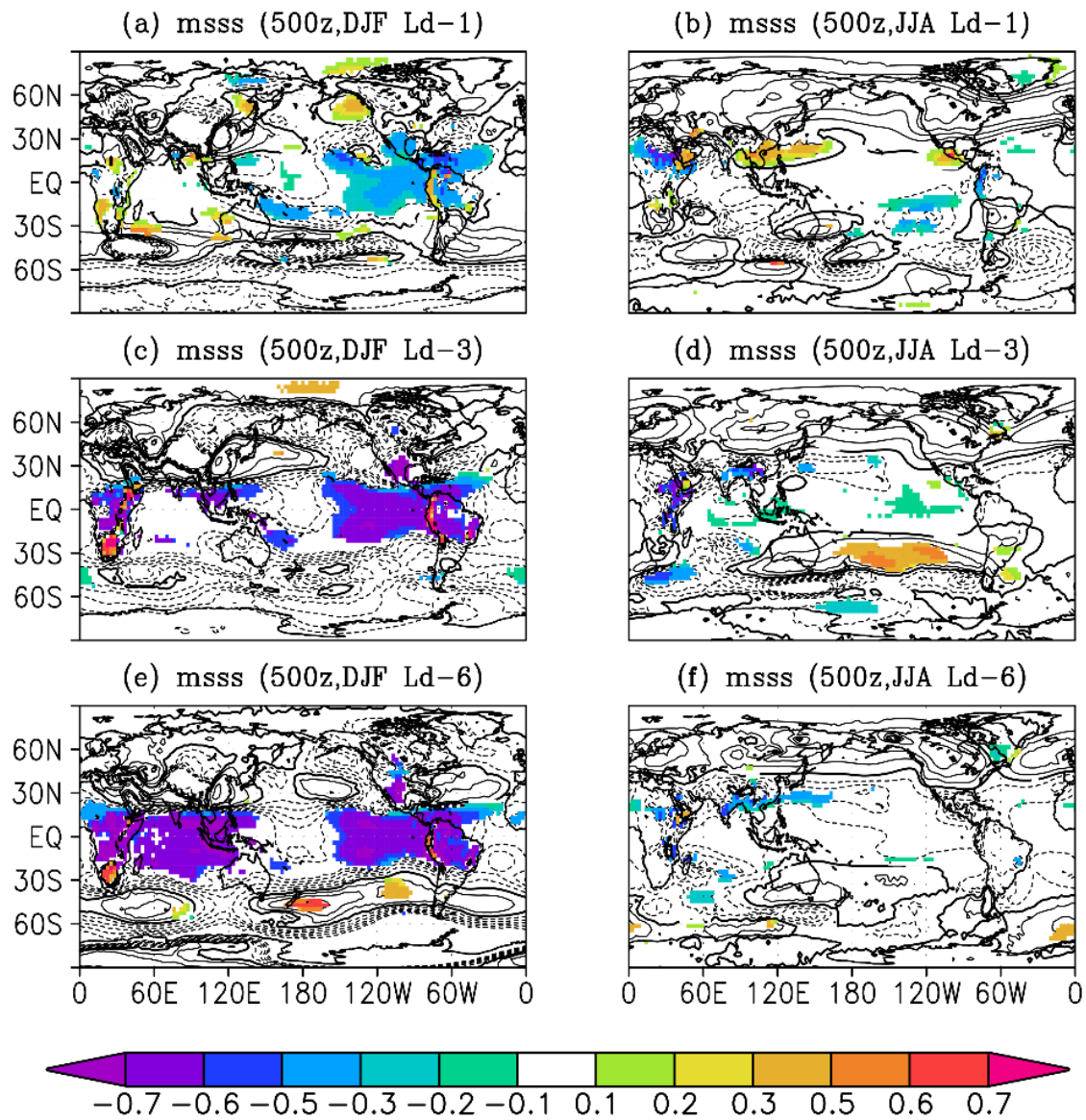


Figure 10. As Figure 9 but for 500hPa GH.

the differences are better explained by model biases which tend to be intensified during neutral ENSO episodes rather than the coupling issue, *per se*.

Furthermore, what has transpired from the comparative analyses of pressure fields is that the differences between the GCMs are a function of both space and time of the year (i.e., seasonality). For example, the CGCM is superior over the Asian monsoon region during the austral winter, but skill deteriorates rapidly with increasing lead-time or for a different season. Although both the CGCM and the AGCM are skillful but not necessarily over the same areas, seasons or even lead-times, combining the forecast from these two models in a multi-model system may further improve on the forecasts.

Further, we examine the implication of variations in the pressure fields of the GCMs on ENSO and equatorial IOD (which are the main climate variability modes and most relevant at the seasonal timescale) and coupling responses (teleconnections). ENSO characteristics are represented using the Southern Oscillation Index (SOI), the mslp difference between Tahiti (17.5°S, 149.5°W) and Darwin (12.5°S, 130.9°E), using the method suggested by *Ropelewski and Jones [1987]*. Likewise, the IOD characteristics are measured using a pressure index with anomalous difference between the mslp in the east and western tropical Indian Ocean [*Saji et al., 1999*]. The indices of these climate drivers are deduced from mslp since previous observational studies show strong association between mslp and SST indices [e.g. *Philander, 1990; Behera and Yamagata, 2003*]. The comparison is based on the November initialized hindcasts since this month coincides with the onset of the seasonal peaks of ENSO although IOD is more active during the austral spring [*Beraki et al., 2014; Zhao and Hendon, 2009*]. By using a Taylor diagram [*Taylor, 2001*], Figure 11 presents skill comparisons by various forecasting methods in predicting ENSO and the equatorial IOD. The skill is represented in the correlation [*Wilks, 2006*] and standard deviation space of the Taylor diagram. The standard deviations are normalized with the corresponding observed standard deviation to

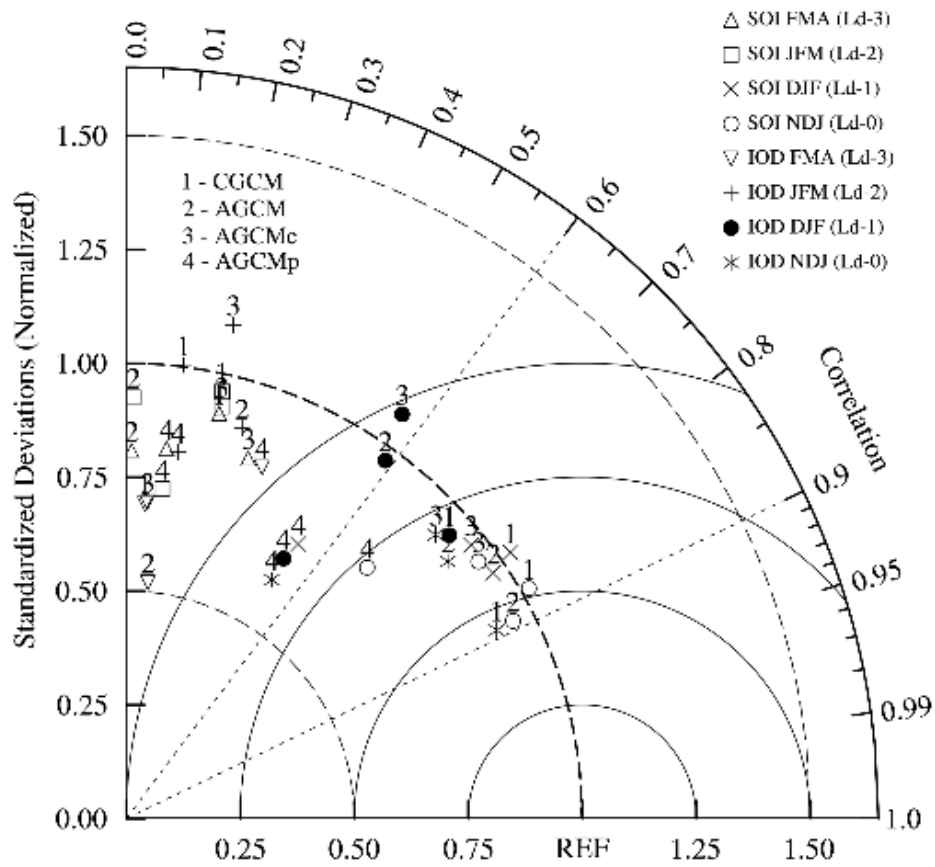


Figure 11. Taylor diagram by CGCM, AGCM, AGCMc and AGCMp (AGCM forced with persisted SST anomalies) based on the ensemble mean from the November initialized hindcasts for the equatorial Indian Ocean Dipole (IOD) and ENSO forecasts (SOI). The standard deviation is normalized by the respective NCEP/DOE. The indices are computed from the respective mslp fields (see text).

facilitate the comparison. The result reveals that differences in the skill and interannual variability between the CGCM and the AGCM at zero-month (November-December-January; NDJ) and one month (DJF) lead-times in predicting ENSO is marginal despite the ability of the GCMs to predict the mslp varying significantly, mainly because the biases across the Pacific Ocean region (between east and west dissect) cancel each other. The CGCM, however, performs better than the AGCM in predicting the equatorial IOD. The result further shows that the CGCM's predictive skill is found to be consistently superior to the AGCM's, a result also reported elsewhere [e.g. *Graham et al., 2005*]. Furthermore, the AGCM's ability is slightly reduced in predicting both ENSO and IOD from the CGCM or AGCM, suggesting that the AGCM benefits from the multi-model SST forcing. At longer lead-times, the GCMs underestimate the observed variability with a sharp skill drop suggesting weakening of the atmospheric response to ocean variations. The predictability of these climate modes, notably ENSO, is much stronger up to several months lead-time when their strength and evolution are measured using SST indices (e.g., Niño 3.4 index) [e.g. *Beraki et al., 2014*] as opposed to mslp derived indices (e.g., SOI) presented here.

The rainfall analysis conducted over various ocean basins along the equatorial Indo-Ocean region and southern Africa sub-continent is presented in Figure 12. During the austral summer, all prediction methods demonstrate nearly similar skills in predicting rainfall and tendencies representing its interannual variability for most regions considered. The exception is that the CGCM performs slightly better than the two AGCM configurations over the equatorial western Pacific at one month lead-time while both the AGCM and the AGCMc perform noticeably better than the CGCM over the equatorial eastern Pacific at three month lead-time. Furthermore, the AGCM simulations manifest a tendency to overestimate the interannual variability over the equatorial Indian Ocean and southern Africa region while the CGCM shows a similar tendency over the equatorial eastern Pacific basin as the lead-time

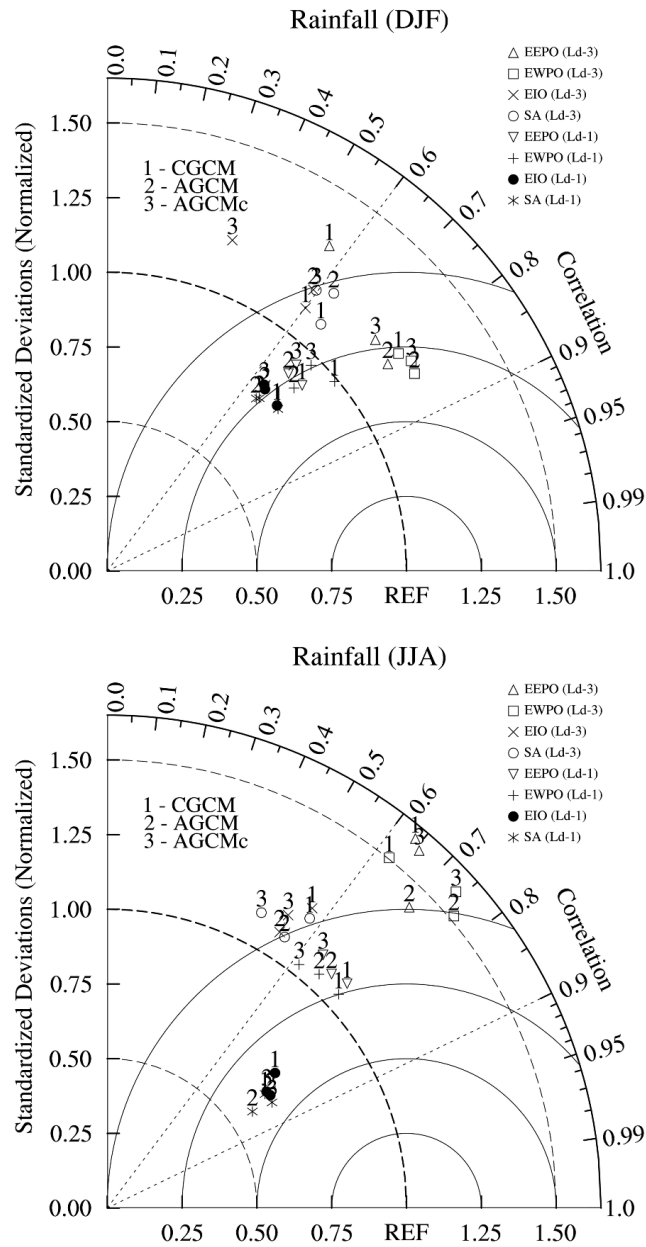


Figure 12. Taylor diagram by one- and two-tiered prediction methods (as shown in the inset) predicting spatially averaged rainfall based on their ensemble means for the southern Africa (SA; 35°S-0 and 0-55°E; masked over the ocean), the equatorial region (20°S-20°N) of Indian Ocean (EIO; 50°E-110°E), western Pacific Ocean (EWPO; 120°E-170°E), and eastern Pacific Ocean (EEPO; 170°E - 60°W). The verification is for the austral summer (top panel) and winter (lower panel) at one and three months lead-times. All standard deviations are normalized by CMAP for the respective basins.

increases. The other difference worth mentioning is that in most instances the AGCM skill is slightly better than the AGCMc and the improvement deepens over the equatorial Indian Ocean at three month lead-time.

The GCMs demonstrate similar levels of skills during the JJA season at one month lead-time as in the case for the DJF season. At three month lead-time, the biggest skill difference between the AGCM and CGCM is found over the eastern and western Pacific region in favor of the AGCM. The AGCM skill is also found to be better than the AGCMc noticeably over the eastern part of the basin. This skill improvement is presumably attributed to the multi-model SST forcing. By and large, rainfall variability over southern Africa and the equatorial Indian Ocean is severely underestimated in all prediction methods with a tendency to be slightly worse in the AGCM. In addition, the interannual rainfall variability is also over or underestimated more during the austral winter than summer season.

The result further indicates that (at least for the austral summer at a one-month lead-time) there is a noticeable similarity in the manner in which the GCMs vary in their skills in the prediction of ENSO (Figure 11), and of the rainfall of most regions (Figure 12). In both cases, the superiority of one model over the other is nearly indistinguishable which suggests ocean-wide atmospheric response to ENSO in both models. The contribution of the equatorial IOD is not clearly manifested on the rainfall predictability of most of the regions considered in the analysis in favor of the CGCM, despite the CGCM's superiority over the AGCM being noticeable in the prediction of IOD. This lack of teleconnection in the GCMs is not clear and is deferred for future work. Observational studies [*e.g. Yang et al. 2010*] report the strong association of IOD with the Asian monsoon during the peak season.

4.2 Comparison based on probabilistic forecasts

In this comparative experiment, the probabilistic scores are calculated from three equiprobable categories, defining below-normal, near-normal and above-normal. The categories are identified from the 33rd and 67th percentiles of the 28 years climatological record. The relative operating characteristic (ROC) area is commonly applied to probabilistic forecasts to measure the ability of a forecasting system to discriminate events such as flood or drought seasons from non-events [Mason and Graham, 1999]. Skillful probabilistic forecast, therefore, possesses higher frequencies of hit rates than false alarms in order to yield the area beneath the ROC curve to be greater than 0.5. The global distributions of the ROC score differences between the CGCM and the AGCM during the austral summer based on the November initialized hindcasts in predicting years of wet and dry conditions are shown in Figure 13. The ROC scores are independently computed against the CMAP rainfall estimates first, and only those scores which are statistically significant at 95% are retained for the comparison, meaning that those probabilistic forecasts which are not better than guessing are omitted. The significance test is conducted using a variant of the Mann-Whitney non-parametric procedure that explicitly accounts for variance adjustment caused by incidents of ties [Mason and Graham, 2002; Wilks, 2006]. From a visual inspection, the CGCM is seemingly doing better over the southern African sub-continent, southern Indian Ocean and Pacific region near the equator. The AGCM, on the other hand, is more successful over the equatorial Indian Ocean off the coast of eastern Africa, over the central and eastern Pacific around 10°N and southern Pacific region off the coast of north-eastern Australia. Broadly speaking, however, the two models are more or less similar in their ability to discriminate below- and above-normal rainfall conditions and most of the differences are as small as 0.1 or 0.15.

According to the global surface temperature skill comparison (Figure 14), the GCMs differ significantly in their ability to differentiate warm or cold episodes from non-events.

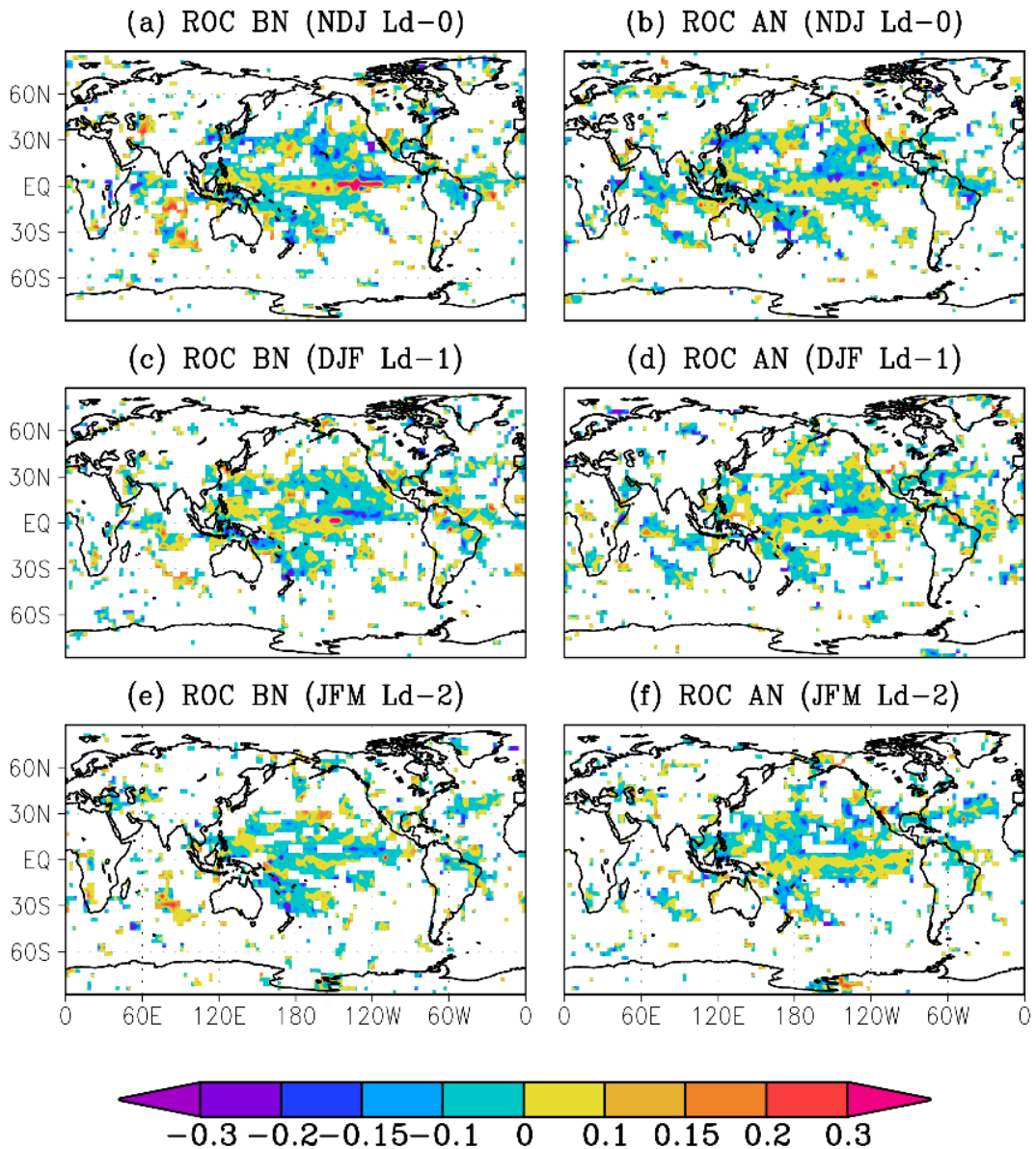


Figure 13. ROC area differences between the CGCM and AGCM for seasonal rainfall totals (mm). The +ve (-ve) scores imply the CGCM (AGCM) is better in discriminating dry (a,c,e) or wet (b,d,f) episodes than the AGCM (CGCM). These differences are computed using the November initialized integrations for various seasons and lead-times as shown in the title of each plot and the skills are independently computed first against the corresponding CMAP estimates. The differences are statistically significant at the 95% level.

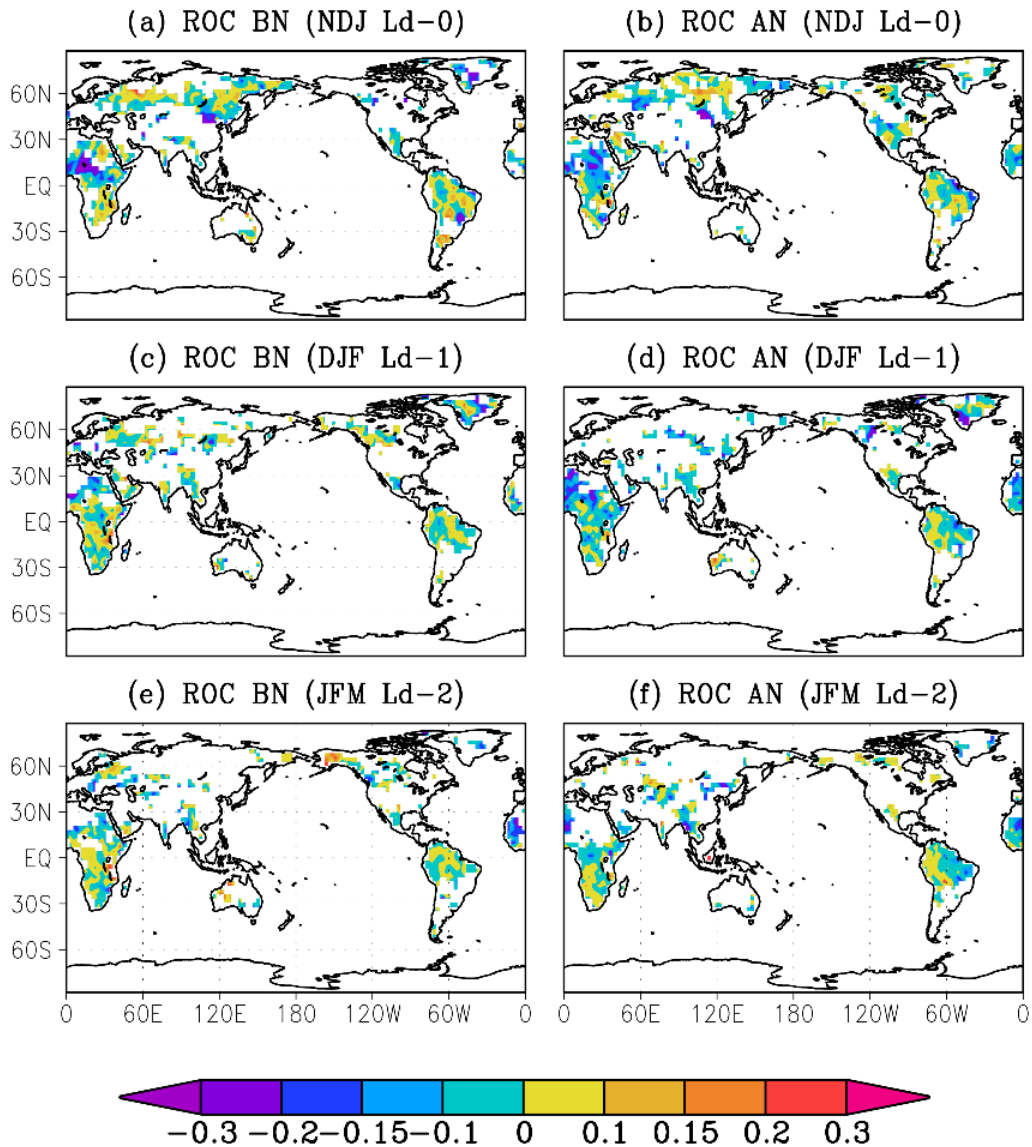


Figure 14. ROC area difference between the CGCM and AGCM for 2m temperatures. The +ve (-ve) scores imply the CGCM (AGCM) is better in discriminating cold (a,c,e) or hot (b,d,f) episodes than the AGCM (CGCM). These differences are computed using the November initialized integrations for various seasons and lead-times as shown in the title of each plot and the skills are independently computed first against the corresponding CRU estimates. The differences are statistically significant at the 95% level.

The skill variations are, however, spatially and seasonally dependent. For instance, during the NDJ season at a zero month lead-time, the AGCM outperforms the CGCM over equatorial Africa for both below- and above-normal temperature conditions (Figure 14a, 14b) with a tendency to persist into the DJF season at a one month lead-time for the upper tercile (Figure 14d). But the condition is changed in favor of the CGCM as the lead-time increases for the lower tercile (Figure 14c, e). Elsewhere, there is an apparent equal distribution of the CGCM and the AGCM predominance.

We extend the comparison further by including various model setup options of the AGCM. The analysis mainly focuses on the seasons surrounding the austral summer, since, as noted earlier, it is an active period of ENSO. Additionally, it is the period when maximum skill is mostly found, particularly at the southern Africa sub-continent at the seasonal timescale. Table 1 shows the ROC area analysis aggregated over different regions by various forecast methods in predicting below- and above-normal surface air temperature conditions. In this comparison, the CGCM is compared with the AGCM with a full ensemble size, reduced ensemble size (AGCM_r), the AGCM forced with the CGCM predicted SSTs (AGCM_c) and persistence (AGCM_p). The last three have the same ensemble size as that of the CGCM. The reduction in ensemble size is made by retaining the AGCM simulations that use multi-model ensemble mean SSTs as lower boundary conditions only. The idea is to explore how the AGCM forecast quality fluctuates by differing them both in terms of ensemble size and SST forcing. The result shows that the AGCM's fidelity in discriminating hot and cold events from non-events is reasonably reduced in the case of AGCM_p and its superiority to the CGCM is noticeably lost in all regions and lead-times considered. The result is consistent to the deterministic skill presented above (Figure 11). The ability of the AGCM to distinguish events from non-events is also slightly reduced with the reduction of

Table 1. Probabilistic skill of the GCMs in predicting cold and warm events as measured using the ROC area aggregated over three different regions^a.

| Lead | Tropics | | | | | Southern Extratropics | | | | | Southern Africa | | | | |
|-------------|---------|------|------|------|------|-----------------------|------|------|------|------|-----------------|------|------|------|------|
| | T1 | T2 | T2r | T2c | T2p | T1 | T2 | T2r | T2c | T2p | T1 | T2 | T2r | T2c | T2p |
| Cold events | | | | | | | | | | | | | | | |
| 0 | 0.66 | 0.67 | 0.65 | 0.63 | 0.68 | 0.59 | 0.56 | 0.55 | 0.52 | 0.56 | 0.62 | 0.60 | 0.59 | 0.59 | 0.60 |
| 1 | 0.69 | 0.68 | 0.67 | 0.65 | 0.65 | 0.60 | 0.60 | 0.60 | 0.56 | 0.57 | 0.69 | 0.66 | 0.63 | 0.66 | 0.60 |
| 2 | 0.70 | 0.70 | 0.69 | 0.68 | 0.64 | 0.59 | 0.61 | 0.60 | 0.59 | 0.54 | 0.77 | 0.73 | 0.72 | 0.73 | 0.62 |
| 3 | 0.73 | 0.72 | 0.71 | 0.70 | 0.62 | 0.60 | 0.62 | 0.62 | 0.63 | 0.52 | 0.77 | 0.74 | 0.72 | 0.74 | 0.59 |
| Warm events | | | | | | | | | | | | | | | |
| 0 | 0.68 | 0.70 | 0.68 | 0.66 | 0.68 | 0.60 | 0.57 | 0.58 | 0.55 | 0.57 | 0.65 | 0.65 | 0.62 | 0.63 | 0.63 |
| 1 | 0.69 | 0.71 | 0.70 | 0.68 | 0.67 | 0.59 | 0.59 | 0.59 | 0.57 | 0.56 | 0.69 | 0.69 | 0.68 | 0.67 | 0.63 |
| 2 | 0.70 | 0.72 | 0.70 | 0.69 | 0.63 | 0.58 | 0.59 | 0.59 | 0.60 | 0.53 | 0.76 | 0.74 | 0.72 | 0.73 | 0.60 |
| 3 | 0.72 | 0.72 | 0.71 | 0.71 | 0.63 | 0.60 | 0.62 | 0.61 | 0.62 | 0.54 | 0.77 | 0.76 | 0.74 | 0.77 | 0.59 |

^aThe analysis is based on the November initialized hindcasts and 0 lead stands for NDJ, 1 for DJF etc. T1, T2, T2r, T2c and T2p represent respectively the CGCM, AGCM, AGCMr, AGCMc and AGCMp. The ensemble sizes used in the analysis for the various forecast strategies are as in Table 1. In this analysis, Tropics and southern Extratropics global zonal belt respectively 20°S – 20°N and 20°S and 20°S-90°S. Southern Africa (SA) as in Figure 12.

the ensemble size (AGCMr) and AGCMc. The latter improvement suggests that it is beneficial to use the multi-model approach to obtain the SST fields to force the AGCM.

In seasonal climate predictions, the forecast quality is better described by virtue of its reliability (calibration) and resolution (sharpness). These measures of skills are commonly practiced to compensate for the potential drawback of the ROC scores in the event where the system is not free of forecast biases. Consequently, we compare the GCMs using the Brier skill score [Murphy, 1973] and the reliability diagrams [Hartmann *et al.* 2002].

The Brier Score (BS) provides a handy measure of accuracy (bias) of probabilistic forecasts aggregated over all forecast probability bins. It has a negative orientation ranging between 0 and 1. In this context, the probabilistic forecast attains perfection when the BS approaches zero. The BS can be decomposed into three terms algebraically i.e., (BS = reliability - resolution + uncertainty) [Murphy, 1973; Wilks, 2006]. A skillful probabilistic forecast therefore attempts to have the lowest possible value and the largest possible value of reliability (B_{rel}) and resolution (B_{res}). Conversely, the uncertainty term (B_{unc}) is independent of the forecast itself and is determined by the inherent circumstance of the observed climatological frequency of the events [see Wilks, 2006].

In this comparative framework, the Brier Skill Score (B_{ss}) is used to measure the relative benefit of one model over the other, i.e., ($BSS = 1 - BS_{CGCM}/BS_{AGCM}$) similar to the MSSS (section 4.1). Likewise, the relative benefits of the reliability and resolution terms are assessed with the same formula, except that the resolution term is normalized by the uncertainty term [Graham *et al.*, 2005].

The CGCM and the AGCM comparison in terms of the BSS, B_{rel} and B_{res} for surface air temperature during the mid-austral summer at one month lead-time is presented in Table 2. The AGCM (with full ensemble size) exhibits a better performance than the CGCM in terms of BSS and B_{rel} for the three regions considered. The maximum benefit of reliability of the

Table 2. The CGCM's benefit relative to various AGCM configurations predicting 2m temperature^a.

| Tercile | Tropics | | | Southern Extratropics | | | Southern Africa | | |
|----------------|------------|------------|------------|-----------------------|-------------|------------|-----------------|-------------|-------------|
| | Bss | Brel | Bres | Bss | Brel | Bres | Bss | Brel | Bres |
| CGCM vs. AGCM | | | | | | | | | |
| Lower | -3.7 | -43.1 | 0.5 | -4.2 | -41.1 | 0.5 | 0.2 | -73.1 | 3.5 |
| Upper | -7.9 | -63.0 | -2.7 | -4.7 | -41.0 | -0.2 | -7.0 | -69.7 | -2.0 |
| CGCM vs. AGCMr | | | | | | | | | |
| Lower | -0.8 | -25.0 | 1.7 | -3.0 | -28.2 | 0.2 | 7.6 | 16.4 | 6.9 |
| Upper | -4.3 | -40.4 | -1.1 | -2.0 | -17.9 | -0.1 | -1.0 | -36.5 | 1.6 |
| CGCM vs. AGCMc | | | | | | | | | |
| Lower | 1.1 | -20.5 | 2.9 | 2.0 | 4.8 | 1.7 | 1.2 | -57.5 | 4.1 |
| Upper | 0.8 | -14.8 | 2.0 | 0.8 | -0.7 | 1.0 | -1.1 | -37.6 | 1.4 |
| CGCM vs. AGCMp | | | | | | | | | |
| Lower | 3.0 | 1.0 | 3.2 | 2.9 | 12.1 | 1.6 | 13.4 | 40.5 | 10.1 |
| Upper | 3.0 | 8.0 | 2.6 | 3.9 | 22.4 | 1.2 | 8.1 | 19.5 | 7.2 |

^aThe relative probabilistic skills are measured using the Brier skill score (Bss) and its algebraic decompositions, i.e., Brier reliability (Brel) and Brier resolution (Bres). Positive CGCM benefits are shown in bold type against various AGCM forecast strategies. The analysis is for DJF at one month lead-time. The CGCM, AGCMr, AGCMc and AGCMp configurations use 10 ensemble members while AGCM uses 30 ensemble members (see text).

AGCM is found in the southern Africa region (73.1% and 69.7% for cold and warm events respectively) followed by the tropics and then the SH extra-tropics. However, in terms of B_{res} , the CGCM (AGCM) tends to outperform the AGCM (CGCM) in predicting cold (hot) events. The AGCMr also attains a better B_{rel} and B_{ss} , with the exception of the below-normal temperature over the southern Africa region, despite the fact that the B_{rel} drops by about 50%. It is noticeable that the CGCM mostly outscores the AGCMc in terms of B_{ss} and B_{res} while the AGCMc performs better than the CGCM in terms of B_{rel} . However, the CGCM has discernibly outperformed persistence (AGCMp) in all Brier terms for all regions.

To gain a deeper insight into their performance differences, the GCMs are further compared using reliability diagrams. The reliability diagram is a graphical tool that is constructed from the computation of the hit rate for the set of forecasts for individual probability bins separately (as opposed to the generalization in the case of the BSS and its decomposition terms), and then plotted against the corresponding forecast probabilities. [Hartmann *et al.* 2002; Wilks, 2006]. The most reliable forecasting system is determined by the extent of its proximity to the diagonal line (perfect reliability).

Figure 15 shows the reliability diagrams for the southern African and Tropical regions. The verification for unusually warm (wet) and cold (dry) events during the austral summer (DJF) at one month lead-time are for the CGCM and for the three cases of the AGCM hindcasts. Also shown is the relative frequency of the use of the forecast bins, which is commonly referred to as the “sharpness diagrams” on the left top corner of each plot both for below- and above-normal conditions. The result shows that the AGCM (forced by the multi-model SST forecasts) and the CGCM both demonstrate similar levels of skill in their ability to detect unusual conditions. Notwithstanding, the CGCM shows relatively more overconfidence than the AGCM at higher probability bins (particularly 0.8), which

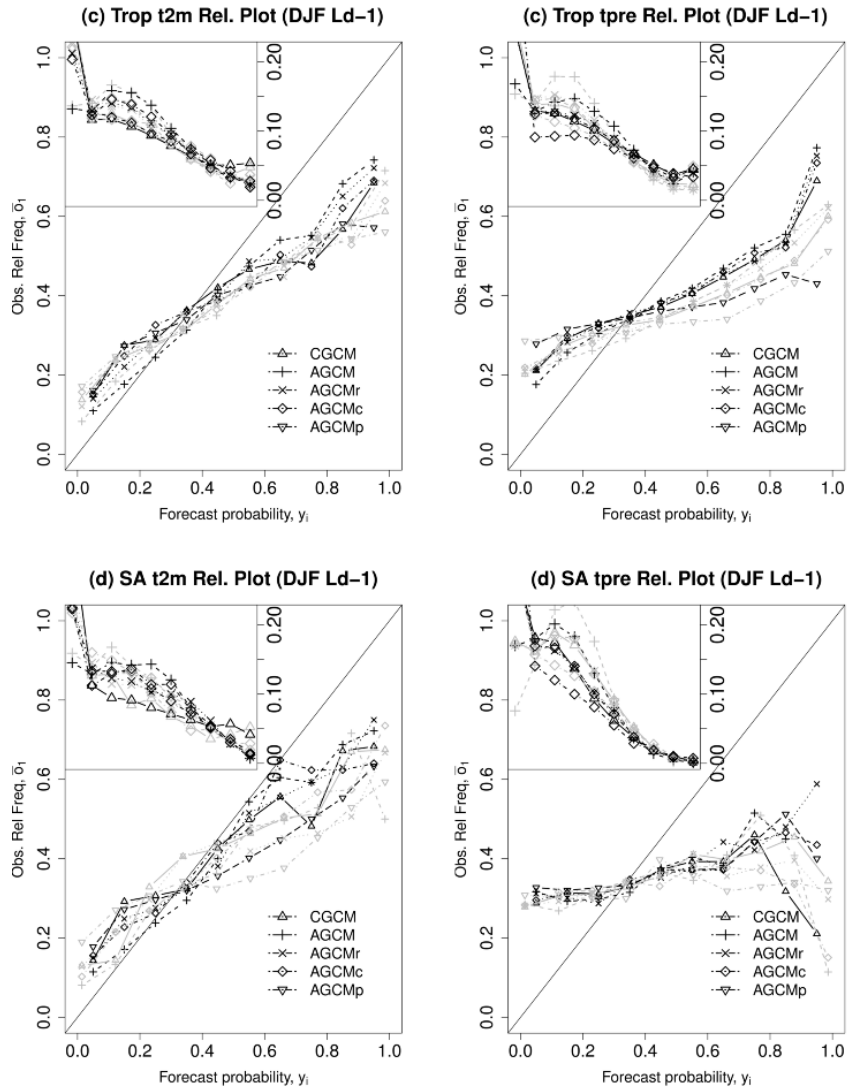


Figure 15. Reliability diagrams by the CGCM (10 ensemble size), AGCM (30 ensembles size), AGCMr (ensemble size reduced to 10), AGCMc (from the SCM predicted SST forced AGCM); ensemble size reduced to 10) and AGCMp (from persisted SST forced AGCM integrations; 10 ensemble size) in predicting below- and above-normal surface air temperature and rainfall conditions during the austral summer seasons (DJF) at one month lead-time for the Tropical region between 20°S and 20°N (top panel) and the Southern African region (bottom panel). The frequency of utilization the different probability bins for both below- and above-normal categories are also shown on the top-left corners of each diagram. The grey and black lines represent cold (dry) and warm (wet) events, respectively.

presumably clarifies the reason why the CGCM is heavily penalized in terms of the B_{ss} and B_{rel} (Table 1).

The result of Figure 15 further indicates that the CGCM and the AGCMc exhibit slightly better sharpness for probabilistic temperature and rainfall predictions respectively, as the sharpness diagrams are flatter when compared to all cases of the other forecast methods. Generally, however, these different forecasting methods tend to fall mostly in the lower or climatological probabilities particularly for rainfall suggesting that the GCMs are more reluctant to issue warnings with higher probabilities.

The reduction of the ensemble size has only caused a minor change in the reliability diagram's shape (i.e. a slight displacement of the curve towards overconfidence when compared to the use of the full ensemble), meaning that the skill drop is too small to alter the circumstance in favor of the CGCM. The AGCMc attains more or less a comparable reliability level to the AGCM (with full or reduced ensemble size), even though its reliability is slightly compromised for probabilistic temperature prediction during the austral summer. Notwithstanding, there is a substantial degradation of skill in favor of the CGCM when the AGCM is forced with persisted SST (AGCMp).

Generally, the result reveals that lack of coupling does not degrade the predictive skill of the AGCM in favor of the CGCM including under a perfect model framework, in which the CGCM SST is used to constrain the AGCM. This result provides evidence of the importance of the role of the predicted SST, notably the multi-model SST, forcing and the oceanic evolution of sea-air interaction in seasonal climate prediction. The study attests to the strongest determining factor in the success of the seasonal prediction being the robustness of the SST information that flows to the GCMs whether prescribed or interactively coupled.

5. Summary and conclusions

The steady increase over recent years in the use of coupled models for seasonal forecasting has been at the expense of uncoupled models owing to the fast development of computational resources and the envisaged advantage of coupled models in representing state-of-the-art seasonal forecasts more realistically. Despite many numerical studies conclusively present evidence in favor of coupled models, a gap still exists whether these models are similar or differ widely in their predictive skill in an operational and hence practical environment. With this in mind, we revisit the subject under a practical model framework, where a multi-model SST anomalies and its uncertainty envelope are used to constrain the atmospheric model. This model comparison study uses the SCM and its atmosphere-only version, which run concurrently at the SAWS for seasonal forecast production in a multi-model environment. Furthermore, the two models are suitably configured in such a manner that the role of coupling on the predictive skill differences is better distinguished. In this experimental framework, the GCMs share a great deal of resemblance in their configuration except for the manner in which the SST information is communicated within the GCMs.

The analysis finds that the two models are able to represent the observed spatial patterns of rainfall and that climatologically, they do not differ strongly in terms of bias distribution both during the SH summer and winter seasons even though the models are somewhat more biased for the latter season. In addition, the comparative analysis reveals that the symmetry and position of the ITCZ and the mid-latitude storm tracks are well represented in both models with a tendency of the CGCM to overestimate the peak of the rainfall distribution in the tropics. This overestimation over the tropics is presumably attributed more to the SST bias than to the air-sea coupling process which is largely minimized in the AGCM with the use of multi-model SST forcing. There are two possible reasons that may substantiate the finding. Firstly, the intensity and distribution of biases are mostly found during the austral

winter period with mostly marginal difference between the GCMs which tends to coincide with the poor predictive skill of ENSO (due to the NH spring barrier). Secondly, the biggest bias differences during the austral summer and winter seasons between the GCMs tend to coincide for the most part with neutral ENSO conditions.

Results from the predictive skill comparisons indicate that most of the differences in the skill of the GCMs arise over the tropical region. Outside the tropics, the superiority of one model over the other is mostly indistinguishable, and the skill levels are also generally lower than over the tropics. The result further indicates that there is a noticeable similarity in the manner in which the GCMs vary in their skills in the prediction of ENSO and rainfall over the equatorial Indo-Pacific basins and the southern Africa sub-continent. In both cases the superiority of one model over the other is mostly indistinguishable and suggests an ocean-wide atmospheric response to ENSO.

In addition, the AGCM's fidelity is drastically reduced in the case of AGCMp to the extent that where the superiority of the CGCM becomes noticeable. The benefit of the AGCM over the CGCM is also slightly reduced with decreasing ensemble size, but not to an extent that can lead to the shift in the superiority balance. Again, this result attests to the conclusion that the role of the multi-model SST forcing is paramount and is the reason why the AGCM and CGCM have comparable levels of forecast skill.

Generally, what has transpired from this comparative experiment is that the GCMs differ widely in their performances, and the issue of the superiority of one model over the other is mostly dependent on space and time (seasonality). One may conclude that the CGCM has the upper hand in the Asian monsoon region during the austral winter. However, the CGCM skill becomes weaker with the increase of lead-time or in a different season. The diversity in their predictive skill as function of space and time may be beneficial in complementing each other in some way.

The modelling work presented here suggests a circumstance under which AGCMs and CGCMs may be able to produce similar levels of skill, notwithstanding the fact that only two such models were considered. At the very least, the study has provided some guidance on how best to optimize an AGCM under circumstances in which limited computational resources only supports the use of AGCMs in an operational forecast environment, a situation commonly found in developing countries such as South Africa. An optimal AGCM configuration, however, depends heavily on skillful SST forecasts, here obtained through a multi-model SST forecast system. These predicted SSTs may be reproduced from a number of CGCMs and such ocean-atmosphere models are therefore essential for skillful seasonal climate predictions when AGCMs are used. A significant amount of work and investment has already gone into AGCM development, but the potential for further improvement of AGCM-based forecasts thus depends to a large extent on the improvement of CGCMs. Nowadays many leading institutions make their seasonal forecast, including SST forecasts, freely available to national and regional centers under the auspices of the World Meteorological Organization (WMO) and so such SST forecasts can be assimilated into AGCM operational forecast systems.

Acknowledgment

The work was supported financially by the Water Research Commission (WRC) and Applied Centre for Climate & Earth Systems Science (ACCESS). The authors are also gratefully appreciative for the CHPC's computational support. Furthermore, computer hardware obtained through the SATREPS (Science and Technology Research Partnership for Sustainable Development) a collaborative project between Japan and South Africa was essential for processing massive model output data. The Max-Planck-Institut für Meteorologie (MPI) has kindly provided the ECHAM4.5 AGCM code. The work was also

impossible without the NCEP reanalysis product. GCMs data used in the study are available for research purpose on request via the SAWS ftp server (ftp://ftp.saws.co.za; email: asmerom.beraki@weathersa.co.za).

References

- Beraki A.F., D.G. DeWitt, W.A. Landman, and C. Olivier, (2014), Dynamical seasonal climate Prediction using an ocean–atmosphere coupled climate model developed in partnership between South Africa and the IRI. *J. Clim.*, 27, 1719-1741.
- Barnston A.G., A. Leetmaa, V. Kousky, R. Livezey, E. O'Lenic, H. Van den Dool, A.J. Wagner, and D Unger, (1999), NCEP Forecasts of the El Niño of 1997-98 and Its U.S. Impacts. *Bull. Am. Meteor. Soc.*, 80: 1829-1852.
- Bengtsson, L., U. Schlese, E. Roeckner, M. Latif, T. Barnett, and N. Graham, (1993), A two-tiered approach to long-range climate forecasting, *Science*, 261, 1026-1029.
- Boville, B.A., and J.W. Hurrell, (1998), Comparison of the Atmospheric Circulations Simulated by the CCM3 and CSM1. *J. Clim.*, 11, 1327-1340.
- Charney, J. G., and J. Shukla, (1981), Predictability of monsoons. Monsoon Dynamics, J. Lighthill and R. P. Pearce, Eds., Cambridge Univ. Press, New York, 99-109.
- Chaudhari, H.S., S. Pokhrel, S. Mohanty, and S.K. Saha, (2013), Seasonal prediction of Indian summer monsoon in NCEP coupled and uncoupled model. *Theor. Appl. Climatol.*, 114, 459–477. doi:10.1007/s00704-013-0854-8
- Colfescu, I., E. K. Schneider, and H. Chen, (2013), Consistency of 20th century sea level pressure trends as simulated by a coupled and uncoupled GCM. *Geophys. Res. Lett.*, 40, 3276–3280, doi:10.1002/grl.50545.
- Copsey D, R.Sutton, and J. R. Knight, (2006), Recent trends in sea level pressure in the Indian Ocean region. *Geophys. Res. Lett.*, 33: L19712, DOI:10.1029/2006GL027175.

- Derber, J., and A. Rosati, (1989), A global oceanic data assimilation system. *J. Phys. Oceanogr.*, 19, 1333–1347.
- DeWitt, D.G., (2005), Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly coupled atmosphere-ocean general circulation model, *Mon. Wea. Rev.*, 133, 2972-2995.
- Doblas-Reyes, F.J., R. Hagedorn, and T.N. Palmer, (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. *Tellus 57A*, 234–252.
- Fu, X.H., B. Wang, and T. Li, (2002), Impacts of air–sea coupling on the simulation of mean Asian summer monsoon in the ECHAM4 model. *Mon Wea. Rev.* 130, 2889–2904
- Goddard, L., S.J. Mason, S.E., Zebiak, C.F., Ropelewski, R. Basher, and M.A. Cane, (2001), Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.*, 21, 1111–1152.
- Graham, R. J., A. D. L. Evans, K. R. Milne, M. S. J. Harrison, and K. B. Robertson, (2000), An assessment of seasonal predictability using atmospheric general circulation models, *Q. J. R. Meteorol. Soc.*, 126, 2211-2240.
- Graham R.J., M. Gordon, P.J., McLean, S. Ineson, M.R., Huddleston, M.K. Davey, A. Brookshaw, and R.T.H. Barnes, (2005), A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction general circulation model. *Tellus 57*, 320–319.
- Hagedorn, R, F.J. Doblas-Reyes, and T.N. Palmer, (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus*, 57A, 219–232.

- Harris I, P. D. Jones, T. J. Osborn, and D. H. Lister, (2014), Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.*, 34 623–642.
- Jha, B., and A. Kumar, (2009), Comparison of the atmospheric response to ENSO in coupled and uncoupled model simulations. *J. Clim.*, 137, 479-487.
- Ji, M., D. W. Behringer, and A. Leetmaa, (1998), An improved coupled model for ENSO prediction and implications for ocean initialization. Part II: The coupled model. *Mon. Wea. Rev.*, 126, 1022–1034.
- Kanamitsu, M., and Coauthors, (2002), NCEP dynamical seasonal forecast system 2000. *Bull. Amer. Meteor. Soc.*, 83, 1019–1037.
- Kirtman, B. P., and Coauthors, (2014), The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, 95, 585-601, doi:10.1175/BAMS-D-12-00050.1.
- Kirtman, B.P., J. Shukla, B. Huang, Z. Zhu, and E.K. Schneider, (1997), Multiseasonal Predictions with a coupled tropical ocean-global atmosphere system. *Mon. Wea. Rev.*, 125, 789-808.
- Komori, N., A. Kuwano-Yoshida, T. Enomoto, H. Sasaki, and W. Ohfuchi, (2008), High-resolution simulation of the global coupled atmosphere-ocean system: Description and preliminary outcomes of CFES (CGCM for the Earth Simulator). High Resolution Numerical Modelling of the Atmosphere and Ocean, K. Hamilton and W. Ohfuchi, Eds., Springer, 241–260.
- Krishnamurti, T.N., C.M. Kishtawal, Z. Zang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, (2000), Multimodel ensemble forecasts for weather and seasonal climate. *J. Clim.*, 13, 4196–4216.

- Kug J.-S., I-S, Kang, and D-H., Choi, (2008), Seasonal climate predictability with Tier-one and Tier-two prediction systems. *Clim. Dyn.*, 31, 403–416. doi:10.1007/s00382-007-0264-7.
- Landman, W.A., and A. Beraki, (2012), Multi-model forecast skill for mid-summer rainfall over southern Africa. *Int. J. Climatol.*, 32: 303–314. DOI: 10.1002/joc.2273.
- Landman W.A., D. DeWitt, D.-E. Lee, A. Beraki, D. Lötter, (2012), Seasonal rainfall prediction skill over South Africa: 1- vs. 2-tiered forecasting systems. *Wea. Forecasting*, 27, 489-501. DOI: 10.1175/WAF-D-11-00078.1.
- L’Heureux, M. L., Collins, D. C., and Hu, Z.-Z. (2012), Linear trends in sea surface temperature of the tropical Pacific Ocean and implications for the El Niño-Southern Oscillation. *Clim. Dyn.*, 1–14. doi:10.1007/s00382-012-1331-2
- Mo, K. C., and M. Ghil, (1987), Statistics and dynamics of persistent anomalies. *J. Atmos. Sci.*, 44, 877–901.
- Molteni, F., L. Ferranti, M. Balmaseda, T. Stockdale, and F. Vitart, (2007), ECMWF seasonal forecast system 3. *CLIVAR Exchange*, 43, 7-9.
- Murphy, A. H., (1988), Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.*, 116, 2417–2424.
- Neelin, J.D., D.S. Battisti, A.C. Hirst, F.-F. Jin, Y. Wakata, T. Yamagata, and S. Zebiak, (1998), ENSO Theory. *J. Geophys. Res.*, 103, 14,261–14,290.
- North, G.R., (1984), Empirical orthogonal functions and normal modes. *J. Atmos. Sci.*, 41, 879–887.
- Pacanowski, R. C., and S. M. Griffes, (1998), MOM 3.0 manual. *NOAA/Geophysical Fluid Dynamics Laboratory Rep.*, 608 pp.
- Palmer, T.N., and D.L.T. Anderson, (1994), The prospects for seasonal forecasting—A review paper. *Quart. J. Roy. Meteor. Soc.* 120, 755–793.

- Palmer, T. N., and Coauthors, (2004), Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), *Bull. Am. Met. Soc.*, DOI: 10.1175/BAMS-85-6-853.
- Philander, S.G.H., (1990), *El Niño, La-Niña and the Southern Oscillation*, Academic Press, pp. 293.
- Roeckner, E., and Coauthors, (1996), Simulation of present-day climate with the ECHAM4 model: Impact of model physics and resolution, *Report No. 93, Max-Planck-Institut für Meteorologie, Hamburg, Germany, 171 pp.*
- Ropelewski, C. F., and P. D. Jones, (1987), An extension of the Tahiti-Darwin Southern Oscillation Index, *Mon. Wea. Rev.*, 115, 2161–2165.
- Shukla, R., and J. Zhu, (2014), Simulations of boreal summer intraseasonal oscillation with CFSv2 over India and western Pacific: role of air-sea coupling. *Atmosphere-Ocean*, 52, 321-330. DOI:10.1080/07055900.2014.939575.
- Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M.P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker, (2014), The NCEP Climate Forecast System Version 2. *J. Clim.* 27, 2185-2208.
- Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H.M. Van den Dool, H.-L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Peña, S. Lord, G. White, W. Ebisuzaki, P. Peng, and P. Xie, (2006), The NCEP Climate Forecast System. *J. Clim.*, 19, 3483-3517.
- Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, (1999), A dipole mode in the tropical Indian Ocean. *Nature*, 401, 360–363.

- Stockdale, T.N., D.L.T Anderson, J.O.S. Alves, and M.A. Balmaseda, (1998), Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, 392, 370–373.
- Tennant, W.J., and B.C., Hewitson, (2002), Intra-seasonal rainfall characteristics and their importance to the seasonal prediction problem. *Int. J. Climatol.*, 22, 1033–1048.
- Troccoli A, M. Harrison, D.L.T. Anderson, S.J. Mason, (2008), *Seasonal Climate: Forecasting and managing risk*. NATO Science Series. Earth and Environmental Sciences Vol 82. Springer: Dordrecht, The Netherlands.
- Wallace, J.M., E.M. Rasmusson, T.P. Mitchell, V.E. Kousky, E.S. Sarachik, and H. von Storch, (1998), On the structure and evolution of ENSO related climate variability in the tropical Pacific: Lessons from TOGA. *J. Geophys. Res.*, 103, 14,241–14,259.
- Wilks, D.S., (2006), *Statistical Methods in the Atmospheric Sciences, 2nd 320 Edition*. Academic Press, pp. 627.
- Xie, P., and P. A. Arkin, (1997), Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bull. Amer. Meteor. Soc.*, 78, 2539–2558.
- Yang, J., Q. Liu, and Z. Liu, (2010), Linking Observations of the Asian Monsoon to the Indian Ocean SST: Possible Roles of Indian Ocean Basin Mode and Dipole Mode. *J. Climate*, 23: 5889–5902. doi: <http://dx.doi.org/10.1175/2010JCLI2962.1>
- Yu J.-Y., and C.R. Mechoso, (1999), A Discussion on the Errors in the Surface Heat Fluxes Simulated by a Coupled GCM. *J. Clim.* 12, 416–426.
- Zhu, J., B. Huang, L. Marx, J. L. Kinter III, M. A. Balmaseda, R-H. Zhang, Z-Z. Hu, (2012), Ensemble ENSO hindcasts initialized from multiple ocean analyses. *Geophys. Res. Lett.* **39**:L09602, DOI: 10.1029/2012GL051503.

Zhu, J., and J. Shukla, (2013), The role of air-sea coupling in seasonal prediction of Asian-Pacific summer monsoon rainfall. *J. Clim.*, 26, 5689-5697, doi: 10.1175/JCLI-D-13-00190.1.