

## Metrics and Diagnostics for Precipitation-Related Processes in Climate Model Short-Range Hindcasts

H.-Y. MA, S. XIE, J. S. BOYLE, S. A. KLEIN, AND Y. ZHANG

*Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California*

(Manuscript received 26 April 2012, in final form 27 July 2012)

### ABSTRACT

In this study, several metrics and diagnostics are proposed and implemented to systematically explore and diagnose climate model biases in short-range hindcasts and quantify how fast hindcast biases approach to climate biases with an emphasis on tropical precipitation and associated moist processes. A series of 6-day hindcasts with NCAR and the U.S. Department of Energy Community Atmosphere Model, version 4 (CAM4) and version 5 (CAM5), were performed and initialized with ECMWF operational analysis every day at 0000 UTC during the Year of Tropical Convection (YOTC). An Atmospheric Model Intercomparison Project (AMIP) type of ensemble climate simulations was also conducted for the same period. The analyses indicate that initial drifts in precipitation and associated moisture processes (“fast processes”) can be identified in the hindcasts, and the biases share great resemblance to those in the climate runs. Comparing to Tropical Rainfall Measuring Mission (TRMM) observations, model hindcasts produce too high a probability of low- to intermediate-intensity precipitation at daily time scales during northern summers, which is consistent with too frequently triggered convection by its deep convection scheme. For intense precipitation events ( $>25 \text{ mm day}^{-1}$ ), however, the model produces a much lower probability partially because the model requires a much higher column relative humidity than observations to produce similar precipitation intensity as indicated by the proposed diagnostics. Regional analysis on precipitation bias in the hindcasts is also performed for two selected locations where most contemporary climate models show the same sign of bias. Based on moist static energy diagnostics, the results suggest that the biases in the moisture and temperature fields near the surface and in the lower and middle troposphere are primarily responsible for precipitation biases. These analyses demonstrate the usefulness of these metrics and diagnostics to diagnose climate model biases.

### 1. Introduction

It has long been a major challenge to simulate precipitation and its related processes correctly in climate models (e.g., Sun et al. 2006; Dai 2006; Lin 2007) because it involves nonlinear interactions of different physical processes over various time and spatial scales, and some of these processes have to be parameterized in current climate models because of insufficient model resolution. As an example to illustrate this issue, Fig. 1 shows the June–August multimodel mean precipitation biases from the World Climate Research Programme (WCRP) phases 3 and 5 of the Coupled Model

Intercomparison Project (CMIP3 and CMIP5)/Atmospheric Model Intercomparison Project (AMIP) experiments (Meehl et al. 2007; Taylor et al. 2012). Significant precipitation biases are present all over the tropics in both CMIP3 and CMIP5 climate models. The patterns and magnitudes of precipitation biases are also similar in both multimodel means.

To understand the causes for such tropical precipitation biases in these AMIP-type climate simulations alone is undoubtedly challenging given the underlying nonlinear feedback processes and compensating errors in the model physics. An alternative and computationally efficient way to diagnose climate errors is through the numerical weather prediction (NWP) technique (Phillips et al. 2004; Rodwell and Palmer 2007; Martin et al. 2010). With a properly and realistically initialized climate model in the forecast mode, one can determine the initial drift from the observations or operational analysis. Assuming that the initial drift due to errors in

Corresponding author address: Hsi-Yen Ma, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Mail Code L-103, 7000 East Avenue, Livermore, CA 94551-0808.  
E-mail: ma21@llnl.gov

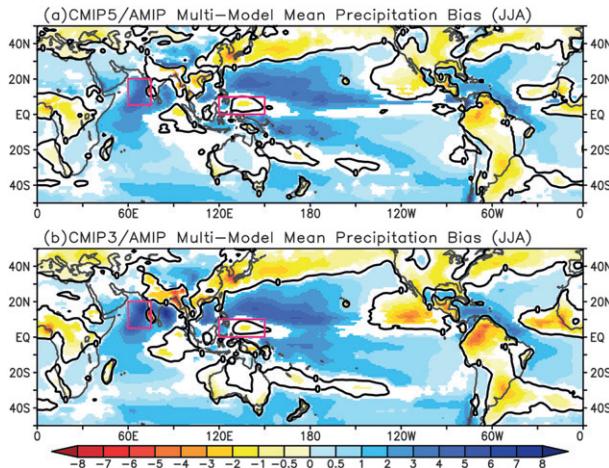


FIG. 1. Biases of June–August multimodel mean precipitation ( $\text{mm day}^{-1}$ ) in reference to GPCP climatology (1980–99) from (a) CMIP5/AMIP and (b) CMIP3/AMIP simulations. Values that are statistically significant at 95% confident level are shaded. The contours indicate zero precipitation. Boxes indicate the regions for analysis in section 4.

the initial conditions is small during the “relatively short” forecast period, a majority of the biases can be attributed to the deficiencies in model parameterizations. Many recent studies have proven the usefulness of such NWP approach for understanding climate model errors and facilitating model parameterization improvements (e.g., Xie et al. 2004; Boyle et al. 2005; Williamson et al. 2005; Klein et al. 2006; Williamson and Olson 2007; Boyle et al. 2008; Williams and Brooks 2008; Xie et al. 2008; Hannay et al. 2009; Martin et al. 2010; Boyle and Klein 2010; Lin et al. 2012; Xie et al. 2012).

To facilitate the use of NWP approach in climate model evaluation and development and to allow for an effective assessment of model performance with both satellite observations and detailed field data, we compiled several metrics and diagnostics in this study, especially for climate model hindcasts on tropical precipitation-related processes. Applying metrics to systematically evaluate climate model performance in simulating mean climatology or variability have become a major focus in climate research community with the availability of WCRP CMIP3 and CMIP5 archives. Some studies focused on the overall performance of model mean climatology or variability (Dai 2006; Gleckler et al. 2008; Pincus et al. 2008; Reichler and Kim 2008; Yokoi et al. 2011), and some studies focused on process-oriented model performance, such as the Madden–Julian oscillation (Kim et al. 2009; Gottschalck et al. 2010; Kim et al. 2011; Sperber and Kim 2012) or monsoon systems (Boo et al. 2011; Wang et al. 2011). In numerical weather forecast centers, a standard set of measures, such as anomaly correlations of geopotential height at 500 hPa (Z500)

over  $20^{\circ}\text{S}$ – $80^{\circ}\text{S}$  and  $20^{\circ}\text{S}$ – $80^{\circ}\text{N}$  or anomaly correlations of zonal winds at 200 hPa (U200) over  $20^{\circ}\text{S}$ – $20^{\circ}\text{N}$ , has been used to routinely assess their model forecast skills at synoptic and intraseasonal time scales (Krishnamurti et al. 2003; Seo et al. 2005).

Different from these commonly used metrics and diagnostics in climate and weather forecast communities, the metrics and diagnostics assembled in the present study are primarily used for systematically exploring model errors and their correspondence between short-term hindcasts and climate integrations, as well as for finding linkage of model biases to particular physical processes. The term “metric” used here is referred to as a scalar quantity of a statistical measure (e.g., pattern correlations or standard deviations) for evaluation of model performance (Gleckler et al. 2008). The term “diagnostic” is a tool to explore why models may produce such biases in the statistical measures. These metrics and diagnostics should provide an objective measure to quantify how fast hindcast errors approach climate errors and to judge the improvement of model simulations from new parameterizations. This is under the premise that improvement of these errors in the hindcast mode is very relevant to the improvement of these errors in the climate mode. Indeed, climate models with better forecast skills tend to perform well in the climate simulations (Hurrell et al. 2010; Martin et al. 2010) since many of the major systematic biases in the climate simulations are associated with “fast processes” (e.g., clouds and precipitation). Furthermore, a standard set of metrics and diagnostics for climate model forecasts is particularly needed because more and more major modeling centers are adapting the concept of “seamless prediction” across weather and climate time scales in developing their future weather forecast models and climate models (Palmer et al. 2008).

The climate model examined in this study is the latest versions of the National Center for Atmospheric Research (NCAR) and U.S. Department of Energy (DOE) Community Atmospheric Model, version 4 (CAM4) and version 5.1 (CAM5). There are significant differences in the physical parameterizations between the two model versions (see section 2 for more details), and both versions are used in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) simulations. Therefore, the evaluation and comparison between these two model versions in their hindcasts provide insights into understanding their AR5 simulations. They are also useful for model development and the hindcast model intercomparison study (i.e., the Transpose-AMIP project: <http://www.transpose-amip.info>) conducted by the international climate community (Williams et al. 2013). The period of study is during

the boreal summer (June–August) of the Year of Tropical Convection (YOTC; May 2008–April 2010). The YOTC project, organized by WCRP and World Weather Research Programme/The Observing System Research and Predictability Experiment (WWRP/THORPEX), established a coordinated effort of observing, modeling, and forecasting tropical convection (Waliser et al. 2012). A goal of YOTC is to improve the ability of current atmospheric models to represent tropical convection and to study its influences on predictability. Several observational datasets and operational analyses, such as National Aeronautics and Space Administration (NASA) A-Train satellite products (Stephens et al. 2002) and European Centre for Medium-Range Weather Forecasts (ECMWF)–YOTC analysis (Waliser et al. 2012), are available for model–data evaluations.

A companion paper by Xie et al. (2012) provided an overview on the correspondence between short- and long-term systematic errors in CAM4 and CAM5. Our focus here is on the introduction of suitable metrics and diagnostics to further explore and diagnose these biases that are related to tropical precipitation and its associated moist processes. The remainder of this manuscript is organized as follows: Section 2 introduces the observational datasets and operational analysis, as well as the CAM. Sections 3 and 4 describe the metrics and diagnostics for hindcast experiments on global tropical and regional precipitation analyses, respectively. Section 5 summarizes our findings and draws conclusions.

## 2. Datasets and model

### a. TRMM, GPCP observations, and the ECMWF–YOTC analysis

Observational rainfall datasets are adopted from multiple Tropical Rainfall Measuring Mission (TRMM) products: 3A12 (microwave imager rainfall estimate), 3A25 (precipitation radar profile estimate), and 3B42 (adjusted merged-infrared precipitation estimate). These three rainfall estimates are obtained from different sensors or algorithms, which allows for some characterization of observational uncertainty. ([http://disc.sci.gsfc.nasa.gov/precipitation/documentation/TRMM\\_README](http://disc.sci.gsfc.nasa.gov/precipitation/documentation/TRMM_README)).

We also use Global Precipitation Climatology Project (GPCP) v2.2 (Adler et al. 2003) for an independent precipitation evaluation (<http://www.esrl.noaa.gov/psd/>). The GPCP data are constructed from over 6000 rain gauge stations, and satellite geostationary and low-orbit infrared, passive microwave, and sounding observations have been merged to estimate monthly rainfall on a  $2.5^{\circ}$  latitude by  $2.5^{\circ}$  longitude global grid from 1979 to the present date.

Other simulated state variables (e.g., temperature or specific humidity) are compared against the operational

analysis datasets from ECMWF–YOTC analysis (Waliser et al. 2012; available online at <http://data-portal.ecmwf.int/data/d/yotc>). The analysis is available at a horizontal resolution of  $0.125^{\circ}$  latitude by  $0.125^{\circ}$  longitude. All these datasets are linearly interpolated into CAM's resolution ( $0.9^{\circ}$  longitude by  $1.25^{\circ}$  latitude), and daily mean fields were used for analyses.

### b. CAM4 and CAM5

In this study, we use both CAM4 (Neale et al. 2013) and CAM5 (<http://www.cesm.ucar.edu/models/cesm1.0/cam/>), which are both used in the IPCC AR5 simulations. Compared to its earlier versions, one important improvement in CAM4 is in its deep convection. The calculation of convective available potential energy (CAPE) was reformulated to include more realistic dilution effects through an explicit representation of entrainment. In addition, convective momentum transport has been included in the parameterization of deep convection. These two changes have resulted in a significant improvement in many aspects of model convection (Neale et al. 2008).

CAM5 is the latest version of CAM and it contains a range of significant enhancements in the representation of physical processes. Almost all of the physical parameterizations in CAM4 have been changed in CAM5, except for the deep convection scheme. This includes 1) a new moist turbulence scheme that explicitly simulates stratocumulus–radiation–turbulence interactions, making it possible to simulate full aerosol indirect effects within stratocumulus (Bretherton and Park 2009); 2) a new shallow convection scheme that uses a realistic plume dilution equation and closure that accurately simulates the spatial distribution of shallow convective activity (Park and Bretherton 2009); 3) a new two-moment cloud microphysics scheme for stratiform clouds (Morrison and Gettelman 2008), which allows ice supersaturation and features activation of aerosols to form cloud drops and ice crystals; and 4) a new radiation scheme, the Rapid Radiative Transfer Model for GCMs (RRTMG), which employs an efficient and accurate correlated- $k$  method for calculating radiative fluxes and heating rates (Iacono et al. 2000; Mlawer et al. 1997).

In this study, we use both CAM4 and CAM5 with their finite volume dynamic core at a horizontal resolution of  $0.9^{\circ}$  longitude by  $1.25^{\circ}$  latitude. In the vertical, CAM4 has 26 levels while CAM5 has 4 additional levels in the boundary layer, which are necessary to run the new boundary and shallow convection schemes.

### c. Model experiments

A series of 6-day hindcasts prescribed with National Oceanic and Atmospheric Administration (NOAA)

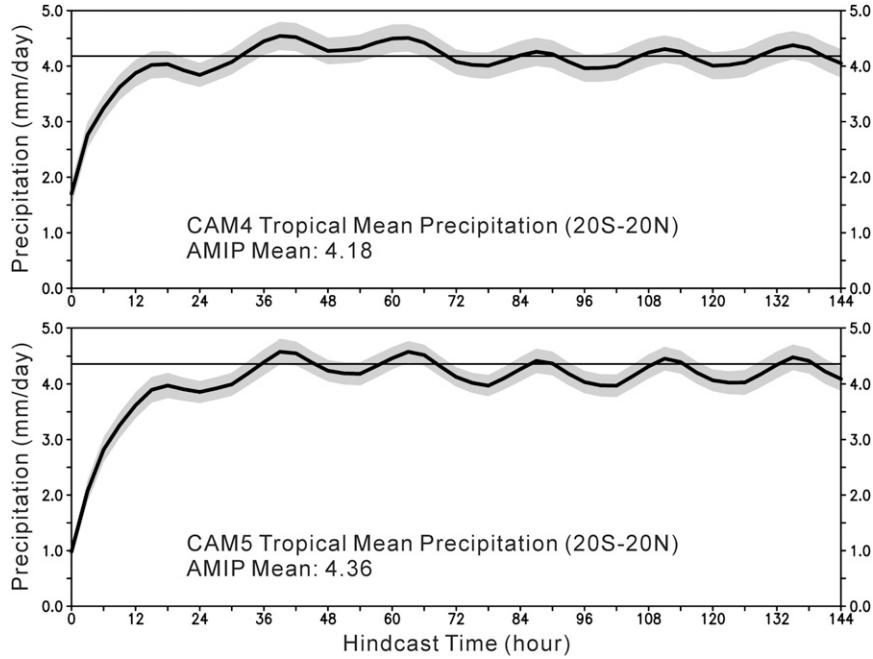


FIG. 2. Ensemble means of tropical June–August mean precipitation ( $\text{mm day}^{-1}$ ) from (top) CAM4 and (bottom) CAM5 hindcasts. The horizontal lines in the plots are their AMIP mean values. Shaded areas are the hindcast standard deviations.

optimum interpolation weekly sea surface temperature (SST; Reynolds et al. 2002) were initialized every day at 0000 UTC using the state variables (winds, temperature, specific humidity, and surface pressure) from the ECMWF analysis for the entire YOTC period under the DOE Cloud-Associated Parameterizations Testbed (CAPT) protocol (Phillips et al. 2004; <http://www-pcmdi.llnl.gov/projects/capt/index.php>). The analysis data are interpolated from the finer-resolution analysis grid ( $0.125^\circ \times 0.125^\circ$ ) to the CAM4/CAM5 grids using the procedures described in Boyle et al. (2005) and Xie et al. (2012). These procedures use a slightly different interpolation approach for each of the dynamic state variables (i.e., horizontal winds, temperature, specific humidity, and surface pressure), along with careful adjustments to account for differences in the representation of the earth's topography between models. Daily hindcast ensembles were calculated using day-1 (0–24 h) to day-6 (120–144 h) hindcasts in order to examine how systematic biases evolve with the hindcast lead time. To compare the statistics of these hindcasts to model simulated climate, we also examine a 3-yr (2008–10) AMIP-type ensemble simulations (three members) prescribed with the same weekly SST. More details regarding CAPT protocol and AMIP experiments are provided by Xie et al. (2012). We will refer to CAPT runs as the short-term hindcasts and AMIP runs as the long-term climate runs.

#### *d. Issues of initial spinup and the hindcasts performance*

Since models are initialized with a “foreign” analysis, effects of initial spinup on the hindcasts are examined through the ensemble means of tropical June–August mean precipitation from CAM4 and CAM5 hindcasts (Fig. 2). The results show that tropical mean precipitation in both CAM4 and CAM5 hindcasts reaches a relative equilibrium state (close to the AMIP means) after  $\sim 24$  h (day 1). This suggests that effects of initial spinup in terms of precipitation have certain impact on day-1 hindcast ensembles but has minimal impact on day-2 or later hindcasts. We can also notice that the spread in the ensemble members are relatively small (standard deviations; gray shaded), especially during the first 24 h. These results are also consistent with those obtained in Williamson et al. (2005) and Boyle et al. (2005). In our analyses, we show all the hindcast ensembles from day 1 to day 6. However, caution should be exercised when interpreting day-1 results.

To further demonstrate both CAM4’s and CAM5’s ability for the hindcast experiments, we plotted in Fig. 3 the anomaly correlations of day-5 Z500 hindcasts over the Northern ( $20^\circ$ – $80^\circ\text{N}$ ) and Southern Hemispheres ( $20^\circ$ – $80^\circ\text{S}$ ), as well as anomaly correlations of day-3 U200 hindcasts over the tropics ( $20^\circ\text{S}$ – $20^\circ\text{N}$ ) for the period June–August 2009. In Fig. 3, both anomaly correlations of

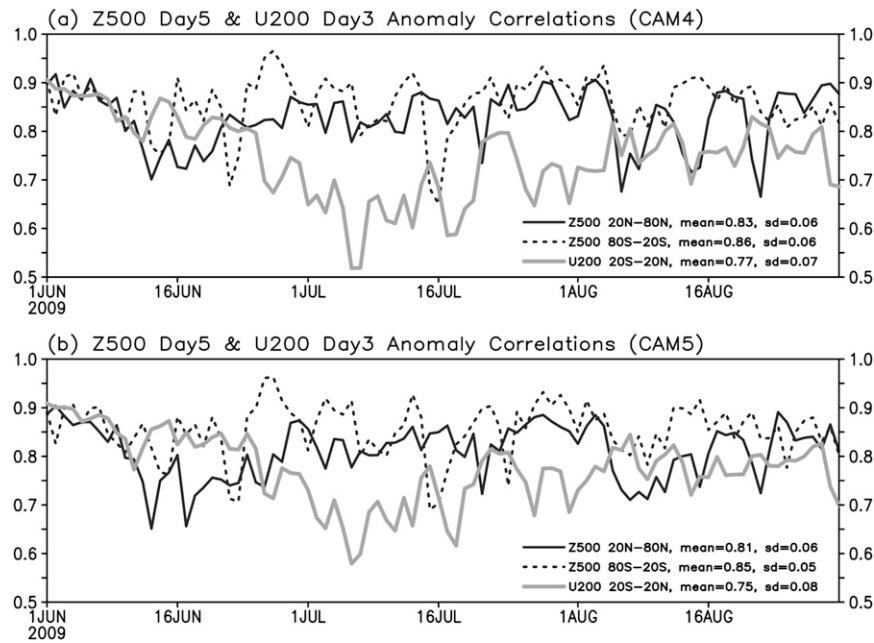


FIG. 3. Anomaly correlations of (a) CAM4 and (b) CAM5 day-5 hindcasts of 500-hPa geopotential height over  $20^{\circ}$ – $80^{\circ}$ N and  $20^{\circ}$ – $80^{\circ}$ S, as well as day-3 hindcasts of 200-hPa zonal winds over  $20^{\circ}$ S– $20^{\circ}$ N.

Z500 for Northern and Southern Hemispheres all show skill scores between 0.7 and 0.95 with mean values of  $\sim 0.85$ . The anomaly correlations of U200 over the tropics show skill scores between 0.5 and 0.95 with mean values of  $\sim 0.75$ . These values are comparable to those obtained from the major NWP models (Yang 2011). Therefore, we are confident that the CAM produces a reasonable large-scale state in which model parameterizations can be evaluated, and we will proceed with our analyses. All our analyses were computed from daily mean fields of June–August of 2008 and 2009 from both hindcasts and AMIP runs.

### 3. Metrics and diagnostics for tropical moist processes

To evaluate precipitation performance and their bias correspondence between short-term hindcasts and long-term climate simulations, our analyses are based on the following metrics: precipitation mean bias, root-mean-square (RMS) errors, pattern correlations, spatial standard deviations, and bias correspondence. We also proposed the following diagnostics: stratiform fraction of precipitation, probability density function (PDF) of daily precipitation intensity, composites of column water vapor (CWV), column relative humidity (CRH; also known as saturation fraction), temperature, and specific humidity profiles as a function of precipitation intensity,

as well as composites of stratiform rainfall fraction as a function of CRH. The physical quantities of CWV and CRH are useful since several previous studies have addressed their strong correspondence to precipitation processes (e.g., Bretherton et al. 2004; Mapes et al. 2009; Neelin et al. 2009; Holloway and Neelin 2010; Sherwood et al. 2010).

Figure 4 displays June–August mean precipitation biases in reference to TRMM 3B42 from CAM4 and CAM5 simulations. Both CAM4 and CAM5 show similar precipitation biases in both CAPT and AMIP runs, and larger biases are mainly present between  $20^{\circ}$ S and  $20^{\circ}$ N. The bias patterns in the CAPT runs in either CAM4 or CAM5 are similar to those in the AMIP runs over most tropical regions, except smaller biases are often seen in the CAPT runs. For example, wet biases are present in the tropical central Pacific Ocean, west of southern Indian peninsula in the Indian Ocean, Central America, northwest corner of South America, and eastern Africa, while dry biases are present near the Maritime Continent, northern and southeastern South America, western tropical Africa, and the tropical eastern Atlantic Ocean. Another striking feature is that precipitation bias patterns in the hindcasts also have high correspondence to those in the multimodel mean biases from the CMIP3 and CMIP5 models (Fig. 1). This feature highly suggests that short-term errors from parameterizations can explain most climate model biases. To consider

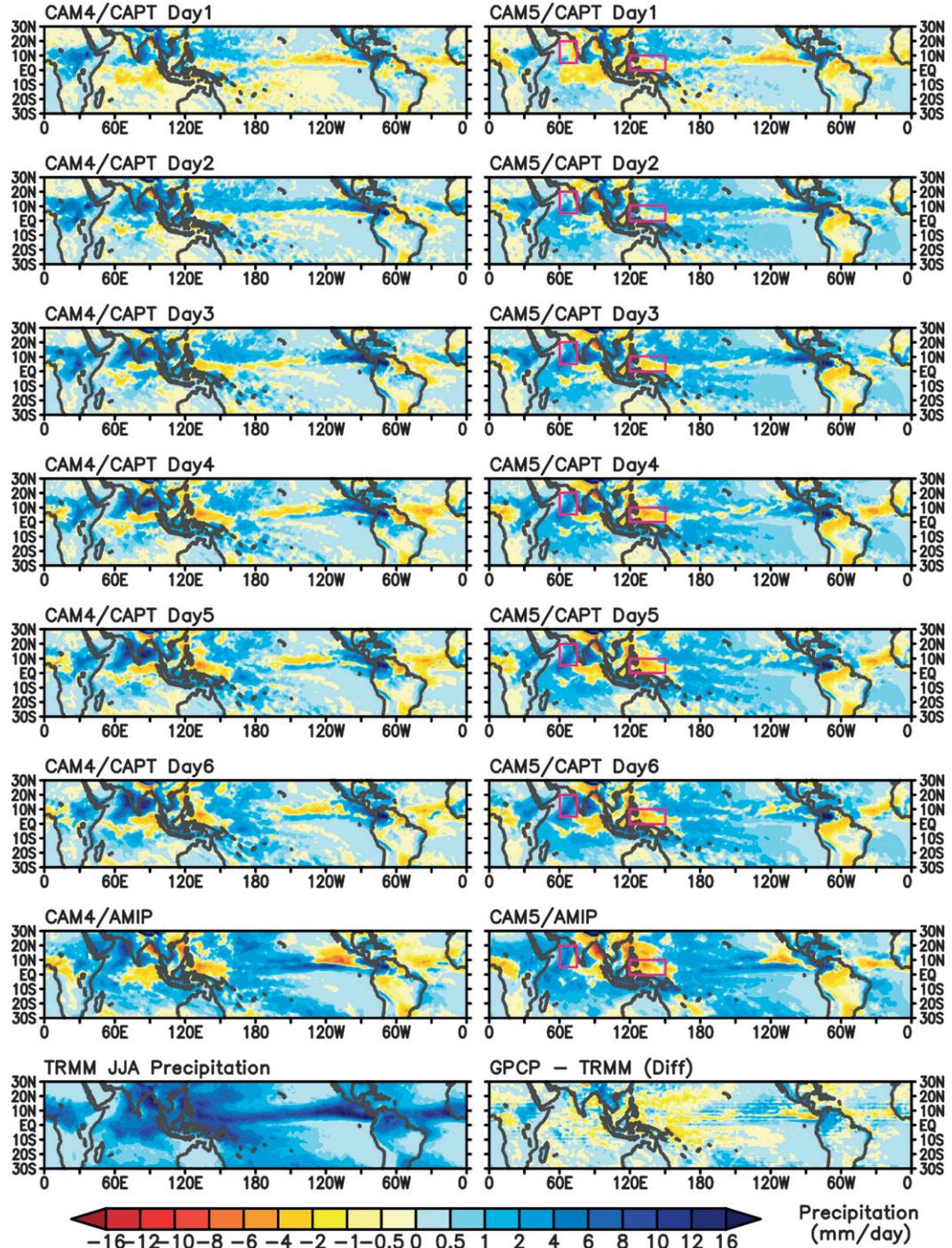


FIG. 4. June–August mean precipitation biases ( $\text{mm day}^{-1}$ ) for CAM4 and CAM5 hindcasts and AMIP simulations in reference to TRMM 3B42. Also plotted is June–August mean precipitation for TRMM and the precipitation difference between GPCP and TRMM. Boxes indicate the regions for analysis in section 4.

observational uncertainty, Fig. 4 also shows the difference of June–August mean precipitation between TRMM 3B42 and GPCP v2.2. The difference is relatively small compared to biases in the simulations, indicating

the robustness of the precipitation bias over many of these regions. The reader can refer to Xie et al. (2012) for a more systematic and quantitative exploration of the correspondence between hindcast and climate errors.

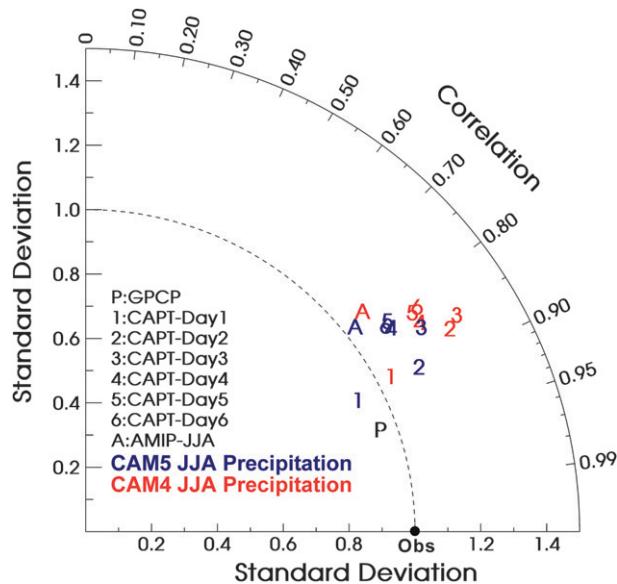


FIG. 5. Pattern statistics of simulated June–August mean precipitation as displayed in a Taylor diagram. The data are analyzed over  $20^{\circ}\text{S}$ – $20^{\circ}\text{N}$ ,  $0^{\circ}$ – $360^{\circ}$ .

To better assess and demonstrate performance skill of these models in simulating tropical precipitation ( $20^{\circ}\text{S}$ – $20^{\circ}\text{N}$ ,  $0^{\circ}$ – $360^{\circ}$ ) and to better understand the model biases, several standard metrics described in Gleckler et al. (2008) are implemented to evaluate the CAPT and AMIP runs. These metrics are presented in Fig. 5 and Tables 1 and 2. Figure 5 summarizes June–August mean precipitation performance in a Taylor diagram (Taylor 2001). The reference dataset “Obs” is from the TRMM 3B42 and is plotted along the  $x$  axis. TRMM and each simulation were normalized by the spatial standard deviations of TRMM such that each field can be shown on the same diagram. Based on Fig. 5 and the statistical metrics in Tables 1 and 2, both CAM5 CAPT and AMIP simulations show slightly better performance in standard deviations and total RMS errors than their correspondent CAM4 experiments, except for the mean biases. The spatial correlations are approximately the same. Both CAM4 and CAM5 simulations show positive mean biases compared to TRMM 3B42 or GPCP v2.2 (Tables 1, 2), and the biases can be as large as almost  $1 \text{ mm day}^{-1}$  in the CAM5 AMIP run. The pattern correlations degraded from day-1 to day-4 hindcasts and then have similar values for day-4–6 CAPT runs in CAM4 or CAM5.

We can better demonstrate the bias correspondence between short-term hindcasts and long-term climate simulations in a Taylor diagram by using biases in the AMIP runs as the reference dataset (Fig. 6). There is a strong correspondence in the pattern statistics between

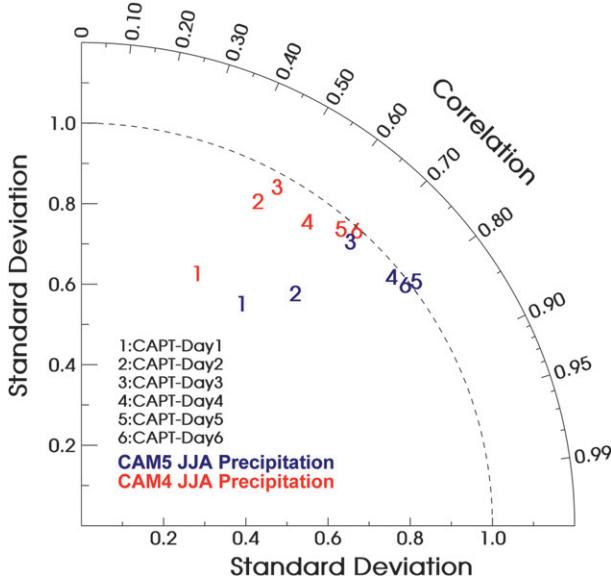
TABLE 1. Statistical metrics for June–August mean global tropical precipitation ( $20^{\circ}\text{S}$ – $20^{\circ}\text{N}$ ) from NCAR CAM4 CAPT and AMIP simulations. The June–August mean tropical precipitation for TRMM 3B42 and GPCP v2.2 are  $3.42$  and  $3.45 \text{ mm day}^{-1}$ , respectively. The reference data are TRMM 3B42.

Simulations	Mean bias ( $\text{mm day}^{-1}$ )	RMS error ( $\text{mm day}^{-1}$ )	Spatial pattern correlation	Normalized spatial std dev ( $\text{mm day}^{-1}$ )
CAPT day 1	0.1	1.74	0.89	1.04
CAPT day 2	0.81	2.41	0.87	1.27
CAPT day 3	0.85	2.58	0.86	1.31
CAPT day 4	0.6	2.39	0.84	1.2
CAPT day 5	0.63	2.5	0.83	1.2
CAPT day 6	0.68	2.57	0.82	1.23
AMIP	0.76	2.61	0.78	1.08

hindcast errors and climate errors (AMIP biases) for precipitation (“fast process” in model physics). The error correlations between these two types of runs are generally larger than 0.5 over the tropics after day-2 hindcasts. The hindcast errors in both CAM4 and 5 gradually evolve toward the AMIP errors with forecast lead time in both correlations and spatial standard deviations. Also, the spatial standard deviations increase from day 2 to day 6 and the magnitude is closer to 1, suggesting the increased bias magnitude in the hindcasts. It is interesting to see that, beyond day-4 hindcasts, the pattern statistics in the later hindcasts are very similar but the day-6 hindcast error does not take all of the way to the AMIP error (with largest error correlations of  $\sim 0.8$  in CAM5). Possible reasons for this can be that the day-6 hindcast is not long enough, and some feedback processes and compensation errors require a longer time scale to develop. Furthermore, the different initial conditions and size of ensembles in the hindcasts and AMIP free runs can also be factors on the correlations. Nevertheless, an error pattern correlation of 0.8 is quite robust to suggest that errors in short-term hindcasts have close resemblance to errors in climate simulations considering these possible reasons here.

TABLE 2. As in Table 1, but for CAM5.

Simulations	Mean bias	RMS error	Spatial pattern correlation	Normalized spatial std dev
CAPT day 1	-0.01	1.58	0.9	0.92
CAPT day 2	0.85	2.01	0.89	1.14
CAPT day 3	0.9	2.43	0.85	1.2
CAPT day 4	0.74	2.38	0.83	1.12
CAPT day 5	0.76	2.46	0.81	1.13
CAPT day 6	0.79	2.43	0.82	1.11
AMIP	0.94	2.53	0.79	1.04



To further understand tropical precipitation biases and related moisture processes, we demonstrated several diagnostics in the following text. One of the useful diagnostics for the precipitation-related processes is the fraction of stratiform precipitation. Correct representation of stratiform-type precipitation in climate models is especially important to determine vertical heating profiles as well as its impact on the general circulation (e.g., Schumacher and Houze 2003; Lin et al. 2004; Dai 2006; Morita et al. 2006; Benedict and Randall 2007; Kim et al. 2009). The separation of precipitation into stratiform and convective contributions in climate models, however, is somewhat ambiguous and can be dependent on model physics and resolutions (e.g., Boyle and Klein 2010). Therefore, one should cautiously attribute an unequivocal physical meaning to such a partition. Nevertheless, Fig. 7 shows the simulated June–August mean fraction of stratiform precipitation averaged over the tropics with regions where seasonal mean total precipitation exceeds 4 mm day<sup>-1</sup>, which is a reasonable good proxy for the threshold of deep convection in the tropics (Chou et al. 2009; Lintner and Neelin 2010). We also plot two observational references from TRMM 3A12 and TRMM 3A25. Schumacher and Houze (2003) concluded ~40% of the tropical precipitation to be stratiform using TRMM 2A23 product, which is based primarily on precipitation radar. In our analysis, TRMM 3A12 shows ~43% of the tropical deep convection to be stratiform, while TRMM 3A25 only shows ~25%. Large

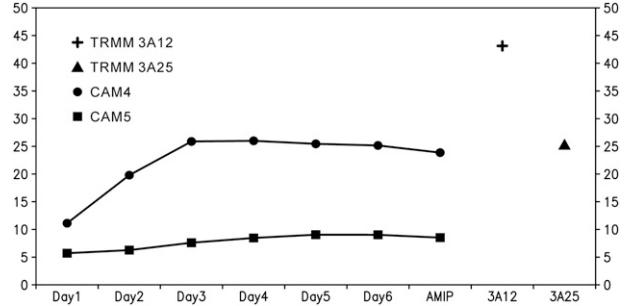


FIG. 7. June–August stratiform rainfall fraction (%) averaged over 20°S–20°N, 0°–360° for TRMM 3A12, TRMM 3A25, and CAM4 and CAM5 hindcasts and AMIP simulations. Only regions where total precipitation exceeds 4 mm day<sup>-1</sup> are calculated.

uncertainty arises because of different retrieval algorithms or instruments used in the observations. For CAM4 AMIP, the fraction is ~24%, and its CAPT runs show significantly smaller fractions in the day-1 hindcast ensemble. The fraction increases significantly from day 1 to day 3 and remains ~24%–25% after the day-3 hindcast. For CAM5, both CAPT and AMIP runs have an average fraction less than 10%. It is not clear what contributes to the difference in the fraction of stratiform precipitation between CAM4 and CAM5, and it warrants further study. Nonetheless, the generally smaller than observed stratiform rainfall fraction simulated by CAM4/CAM5 implies the dominant role played by deep convection, which might be too active. It is also interesting to note that the stratiform fraction in the hindcast runs does not change much after day 3, suggesting the quick model adjustment from the initial conditions.

Another useful diagnostics for tropical precipitation evaluation is the PDF of daily precipitation intensity (Fig. 8). The PDFs are obtained by computing the fraction of data points for each precipitation bin over the entire tropical band (20°S–20°N) of June–August of 2008 and 2009 (182 days). The bin size is 10<sup>0.1</sup> mm day<sup>-1</sup> on the log<sub>10</sub> scale. In Fig. 8, the precipitation intensity PDF on the logarithmic axis from TRMM 3B42 shows a negative skewness with the large percentage of precipitation ~2–4 mm day<sup>-1</sup>. Both CAM4 and CAM5 PDFs, however, show strong bimodal distribution. In CAM4, one peak is ~0.2 mm day<sup>-1</sup> and another is ~4 mm day<sup>-1</sup>. In CAM5, one peak is ~1 mm day<sup>-1</sup> and another is ~7 mm day<sup>-1</sup>, suggesting a shift of probability toward higher precipitation intensity and better simulations of light rain in CAM5. By further examining the correspondence between hindcasts and AMIP runs, we find that, in CAM4, day-1 and day-2 hindcasts show slightly larger probability comparing to other hindcasts for both peaks but smaller probability between 0.5 and 4 mm day<sup>-1</sup>.

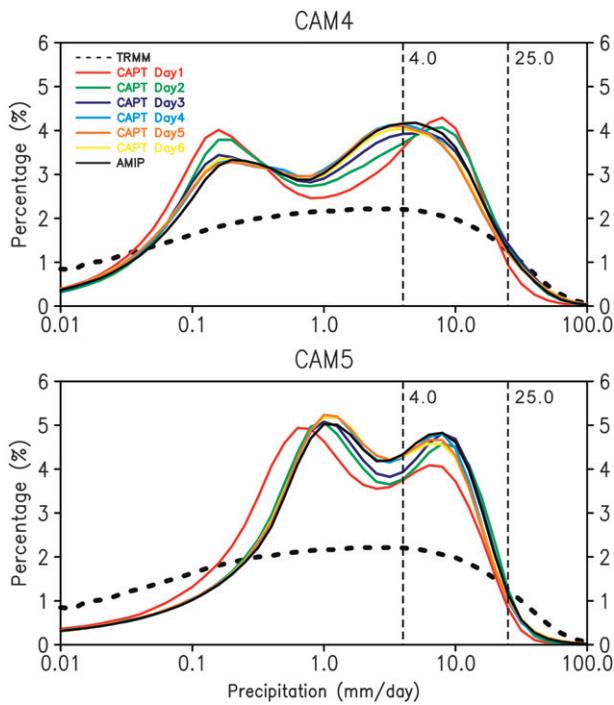


FIG. 8. Daily precipitation PDF for TRMM 3B42, as well as for CAM4 and CAM5 hindcasts and AMIP simulations. The bin size is  $10^{0.1}$  mm day $^{-1}$  on the log<sub>10</sub> scale. The data are analyzed for June–August of 2008 and 2009 and over 20°S–20°N, 0°–360°.

Beyond day 3, the overall structures in the later hindcasts or AMIP ensembles are very similar to each other. Similar correspondence feature is also present in CAM5 simulations, except day-1 hindcasts show smaller probability between 1 and 10 mm day $^{-1}$ . In general, CAM4 has much larger percentage than TRMM for precipitation intensity between 0.04 and 25 mm day $^{-1}$ , and CAM5 has a much larger percentage for precipitation intensity between 0.25 and 25 mm day $^{-1}$ . Furthermore, both CAM4 and CAM5 show lower percentage than TRMM with precipitation intensity larger than 25 mm day $^{-1}$ , although CAM4 has better PDF than CAM5. This indicates that both versions of CAM underestimate or are less likely to produce extreme precipitation events. Nevertheless, short-term hindcast biases again show stronger correspondence to climate biases in precipitation PDF.

To better understand the precipitation PDF bias, we applied two diagnostics: CWV (kg m $^{-2}$ ) and CRH (%). Many observational studies have shown a strong relationship between precipitation intensity and CWV or CRH for deep convection (e.g., Bretherton et al. 2004; Mapes et al. 2009; Neelin et al. 2009; Holloway and Neelin 2010; Sherwood et al. 2010). Based on Bretherton et al. (2004), the CRH is defined as

$$\text{CRH} = \frac{\text{CWV}}{\text{CWV}^*} \times 100\%, \quad (1)$$

where CWV\* is the saturated CWV calculated from the vertically integral of saturated specific humidity.

We plotted in Fig. 9 the CRH and CWV against precipitation intensity in logarithmic scale. Figure 9 shows that the CRH and CWV from ECMWF have a fairly good log-linear relationship with TRMM precipitation intensity, especially for precipitation that exceeds 1 mm day $^{-1}$ . For the same precipitation intensity, both CAM4 and CAM5 generally show larger mean CRH than ECMWF, especially for precipitation intensity larger than 4 mm day $^{-1}$  (exceeds their standard deviations). This indicates that both model versions, especially CAM5, require much higher CRH to generate stronger precipitation events. For CWV, both CAM4 and CAM5 have slightly larger (smaller) mean CWV than ECMWF for precipitation intensity larger (smaller) than 10 mm day $^{-1}$ , although the difference is within one standard deviation range.

Figure 10 further shows the composites of stratiform rainfall fraction against the CRH from CAM4 and CAM5 hindcasts and AMIP runs. The bin size for CRH is 1%. In CAM4, both hindcasts and AMIP runs have similar magnitude. In CAM5, the AMIP run has much a smaller stratiform rainfall fraction between 30% and 80% of CRH. Both models also tend to produce lower fraction between 40% and 80% of CRH. Based on the information from Figs. 8 and 9, we found that, for CRH between 70% and 85%, which is roughly corresponding to precipitation intensity between 4 and 25 mm day $^{-1}$ , the stratiform rainfall fraction is lowest around 20%–30% in CAM4 and 10% in CAM5. These results suggest that the precipitation is largely dominant by cumulus precipitation between 4 and 25 mm day $^{-1}$ . A well-known problem for CAM's deep convective scheme (Zhang and McFarlane 1995) is that the scheme is too often triggered (e.g., Xie and Zhang 2000) such that the atmospheric instability is quickly released. CAPE is frequently consumed, and column moisture is also overly consumed by the precipitation process, leading to biased low specific humidity in the middle and lower levels (shown in Fig. 12). This is consistent with the precipitation PDFs shown in Fig. 8 that precipitation intensity between 4 and 25 mm day $^{-1}$  occurs more frequently than in TRMM. On the other hand, the CRH requirement for extreme precipitation events in the model is much higher than that in the analysis data. This is also consistent with the smaller probability of precipitation events in the simulations than in the TRMM in Fig. 8, as the moisture in the middle and lower levels are overly consumed by the too active deep

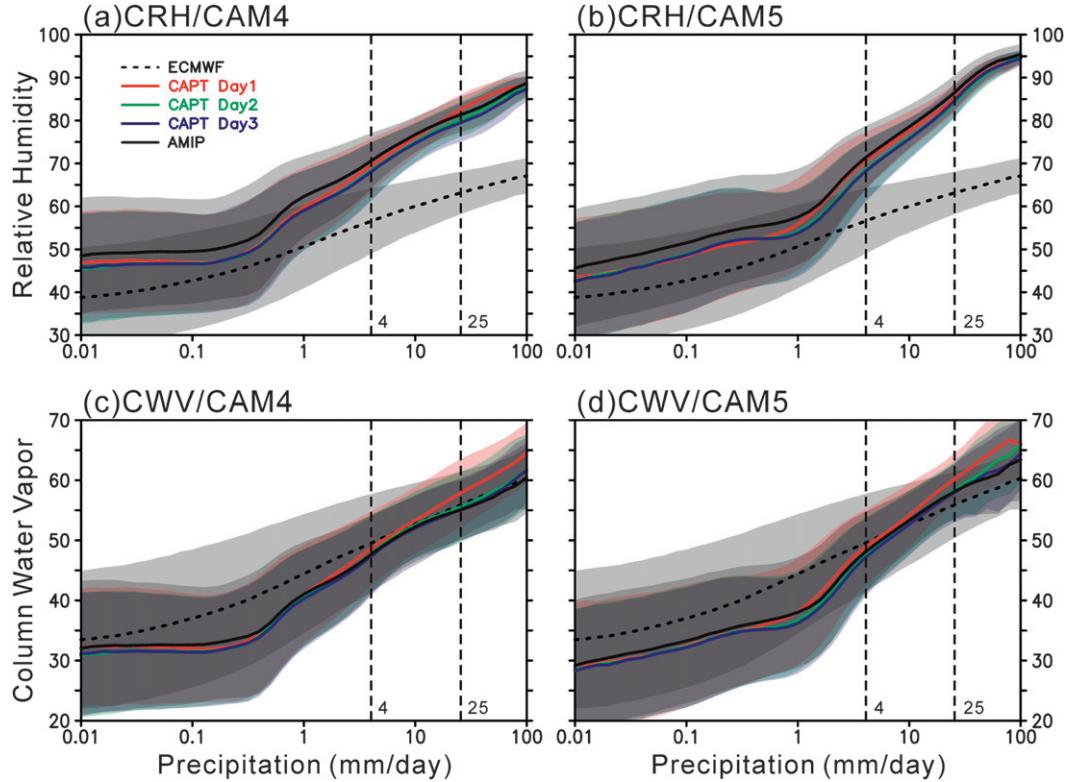


FIG. 9. Composites of daily (a),(b) column relative humidity, and (c),(d) column water vapor as a function of precipitation intensity from ECMWF–YOTC analysis/TRMM 3B42, as well as from CAM4 and CAM5 hindcasts and AMIP simulations. The bin size is  $10^{0.1}$   $\text{mm day}^{-1}$  on the  $\log_{10}$  scale. The data are analyzed for June–August of 2008 and 2009 and over  $20^\circ\text{S}$ – $20^\circ\text{N}$ ,  $0^\circ$ – $360^\circ$ . Shaded areas are their standard deviations.

convection and therefore the high CRH is harder to achieve in the model simulations.

To evaluate the interactions between precipitation and large-scale state variables, we further examined the composited vertical profiles of temperature (Fig. 11) and specific humidity (Fig. 12) diagnostics. The biases were calculated in reference to ECMWF analysis. For precipitation intensity less than  $4 \text{ mm day}^{-1}$ , there are large temperature biases below 800 hPa in both CAM4 and CAM5, but no significant bias tendency is present in the hindcasts as the hindcast lead time increases. For precipitation intensity larger than  $4 \text{ mm day}^{-1}$ , a common cold bias feature is present between 800 and 600 hPa, and a warm bias feature is present below 800 hPa and between 600 and 200 hPa in both CAM4 and CAM5. There is also a cooling tendency between 800 and 500 hPa in the hindcasts from day 1 to day 3 for both CAM4 and CAM5. For specific humidity profiles (Fig. 12), a common dry bias feature above 900 hPa is present for precipitation intensity smaller than  $4 \text{ mm day}^{-1}$  and a wet bias feature of the entire troposphere is present for precipitation intensity larger than  $4 \text{ mm day}^{-1}$ . This is consistent with the CWV features in Fig. 9. Also,

there is a drying tendency for precipitation intensity larger than  $1 \text{ mm day}^{-1}$  between 800 and 500 hPa in the hindcasts from day 1 to day 3 for both versions of model. The drying tendency is consistent with the conversion of moisture to rainfall indicated by the increased probability of precipitation intensity due to deep convection mentioned above. However, there is no clear warming tendency associated with cumulus convection-induced subsidence in the temperature composites. Instead, a cold bias and a cooling tendency in the hindcasts are present in the middle troposphere. This indicates that the biases come from other processes such as evaporation/melting of precipitation, radiative cooling due to biases in clouds, or biases in the model dynamics such as horizontal and vertical advection of temperature.

#### 4. Diagnostics for regional precipitation biases

In the section, we further demonstrate our diagnostics with an emphasis on the analysis of regional precipitation biases. We selected two regions where similar bias pattern and signs are present in both CAM4 and

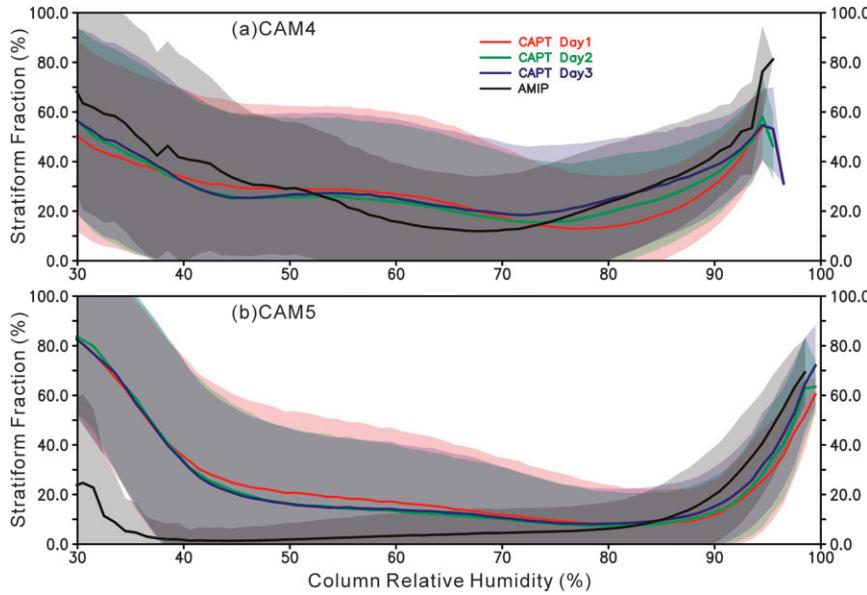


FIG. 10. Composites of daily stratiform rainfall fraction as a function of column relative humidity from (a) CAM4 and (b) CAM5 hindcasts and AMIP simulations. The bin size for column relative humidity is 1%. The data are analyzed for June–August of 2008 and 2009 and over 20°S–20°N, 0°–360°. Shaded areas are their standard deviations.

CAM5 AMIP simulations (boxes in Fig. 4). In addition, both CAM4 and CAM5 hindcasts all picked up such bias patterns, and the bias magnitude generally increases with forecast lead time. This type of analysis over regional scale especially allows for examining the development of bias in the short-term hindcasts before various feedback processes and compensation errors set in. The first region is over the northwestern Pacific warm pool between 120° and 150°E and between 0° and 10°N, where a systematic dry bias is present in both versions of the models. The second region is over the Indian Ocean just near southwest of the Indian peninsula between 60° and 75°E and between 5° and 20°N, where a systematic wet bias is present. Similar bias patterns for these two regions are also presented in the CMIP3/AMIP and CMIP5/AMIP multimodel mean bias (Fig. 1) indicating a common and systematic bias feature in the contemporary AGCMs.

For the regional diagnostics, we computed profiles of large-scale heat and moisture budget residual as represented by  $Q_1$ , and  $Q_2$  (Yanai et al. 1973) and of moist static energy (MSE) from daily mean fields of June–August of 2008 and 2009 based on the following three equations [Eqs. (2)–(4)].

Here,  $Q_1$  and  $Q_2$  are defined as

$$Q_1 \equiv c_p \left( \frac{p}{p_0} \right)^\kappa \left( \frac{\partial \bar{\theta}}{\partial t} + \bar{\mathbf{v}} \cdot \nabla \bar{\theta} + \bar{\omega} \frac{\partial \bar{\theta}}{\partial p} \right) \quad \text{and} \quad (2)$$

$$Q_2 \equiv -L \left( \frac{\partial \bar{q}}{\partial t} + \bar{\mathbf{v}} \cdot \nabla \bar{q} + \bar{\omega} \frac{\partial \bar{q}}{\partial p} \right), \quad (3)$$

where  $c_p$  is specific heat capacity of dry air at constant pressure,  $\theta$  is potential temperature,  $q$  is specific humidity,  $\bar{\mathbf{v}}$  is horizontal velocity,  $\bar{\omega}$  is vertical velocity in the pressure coordinate  $p$ ,  $L$  is latent heat of vaporization,  $\kappa = R \times c_p^{-1}$  with  $R$  being the gas constant, and  $p_0 = 1000$  mb. Overbars represent the running horizontal average over a large-scale area.

The MSE  $h$  is defined as

$$h = c_p T + L_v q + gz, \quad (4)$$

where  $T$  is temperature,  $g$  is gravity,  $z$  is geopotential height, and  $L_v$  is latent heating of vaporization. For saturated MSE (MSE\*), the saturated specific humidity  $q^*$  is used for Eq. (4).

MSE profiles indicate the stability of atmospheric columns based on the vertical distributions of temperature and moisture. Furthermore, correct representation of precipitation process at a given grid requires not only the correct rainfall values at the surface but the associated vertical heating  $Q_1$  and moistening  $Q_2$  profiles. We only performed the analysis over the two selected regions from CAM5 simulations, but similar techniques can be applied to other regions or time scales and with different model simulations. The

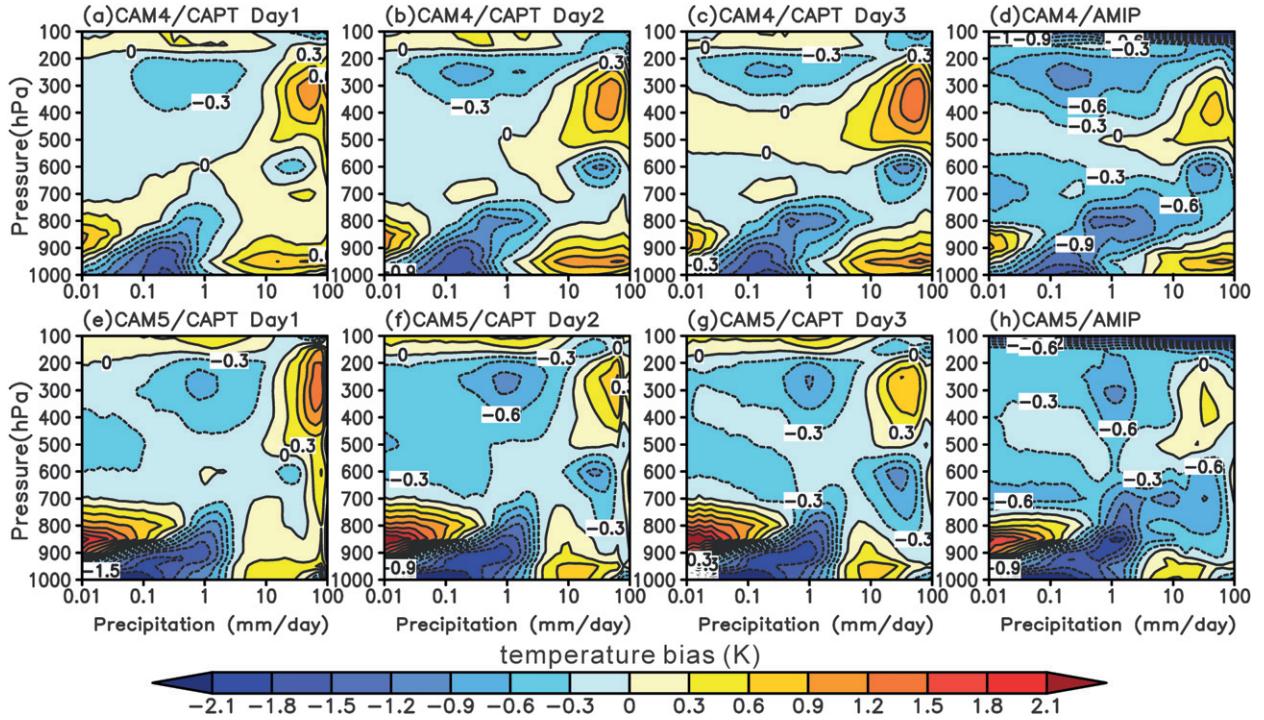


FIG. 11. Biases of temperature profile (K) composites as a function of precipitation intensity from (a)–(d) CAM4 and (e)–(h) CAM5 hindcasts and AMIP simulations in reference to ECMWF–YOTC analysis/TRMM 3B42. The bin size is  $10^{0.1}$   $\text{mm day}^{-1}$  on the  $\log_{10}$  scale. The data are analyzed for June–August of 2008 and 2009 and over  $20^\circ\text{S}$ – $20^\circ\text{N}$ ,  $0^\circ$ – $360^\circ$ .

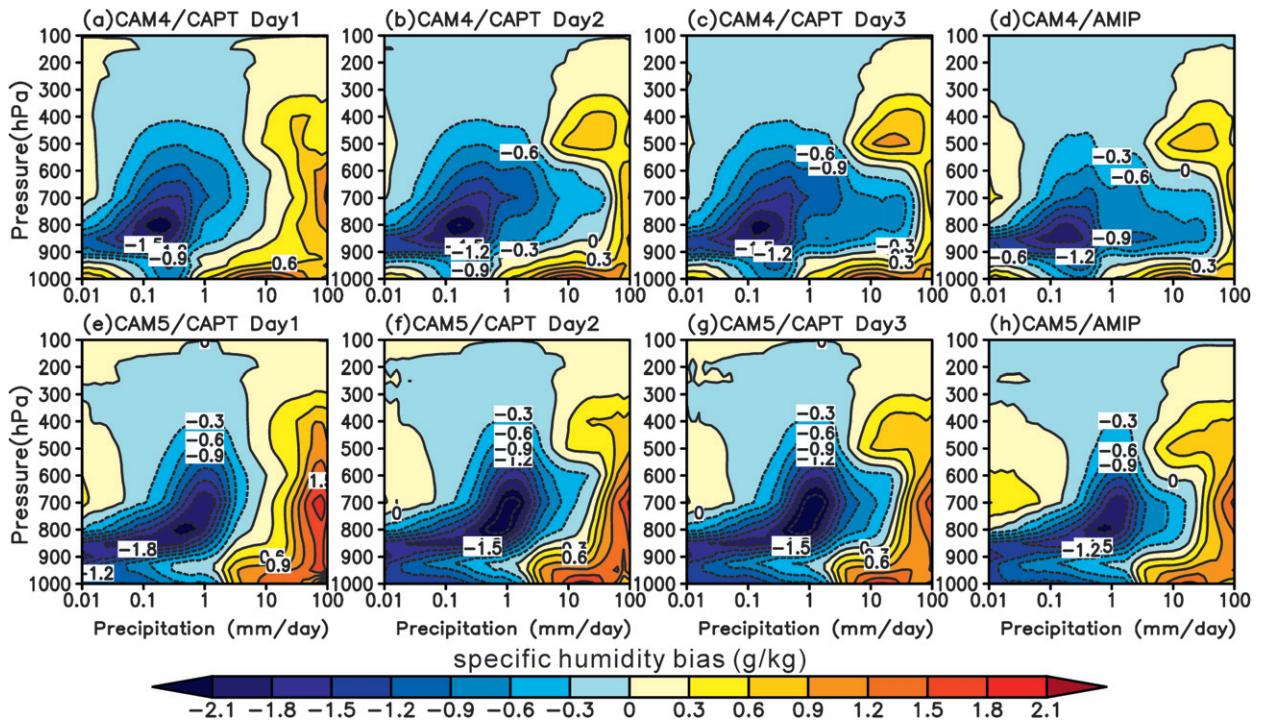
purpose of these diagnostics is to indicate the possible directions that a climate model may follow in producing mean climate biases as illustrated by the hindcast technique. All the calculations are done by applying the above equations with the state variables (winds, vertical velocity, temperature, and specific humidity) from ECMWF–YOTC operational analysis. The  $Q_1$ ,  $Q_2$ , and MSE values for models are also calculated in the same way such that they are consistent and can be compared with ECMWF values.

#### a. Dry bias over the tropical northwestern Pacific warm pool

We first examine the dry bias over the tropical northwestern Pacific warm pool ( $0^\circ$ – $10^\circ\text{N}$ ,  $120^\circ$ – $150^\circ\text{E}$ ). Figure 13 shows CAM5 precipitation mean bias averaged over this region. After day 1, precipitation decreases between days 2 and 5 in the hindcasts. Figure 14 shows the ECMWF June–August mean vertical profiles of  $Q_1$  and  $Q_2$  averaged over the selected domain. The units have been converted into  $\text{K day}^{-1}$  by multiplying by  $c_p^{-1}$ . The  $Q_1$  and  $Q_2$  profiles both show positive heating rates with heating maxima of  $\sim 4$  and  $2.5 \text{ K day}^{-1}$  at 500 hPa, respectively. Such heating and drying profiles are typical warming and drying effects associated with deep convection-induced large-scale subsidence, as well

as detrainment of water vapor and re-evaporation of detrained condensate (e.g., Yanai et al. 1973; Luo and Yanai 1984; Yanai and Tomita 1998). Figure 14 also shows the  $Q_1$  and  $Q_2$  biases from the simulations in reference to the ECMWF analysis. Both simulated  $Q_1$  and  $Q_2$  biases show anomalous diabatic cooling (less warming) and moistening (less drying) from day-2 to day-6 hindcasts as well as the AMIP runs. The anomalous cooling in  $Q_1$  is present almost in the entire troposphere, while the anomalous moistening in  $Q_2$  is mostly confined below 500 hPa. There is also a cooling and moistening tendency below 500 hPa between day-2 and day-4 hindcasts. The biases in  $Q_1$  and  $Q_2$  profiles are consistent with precipitation biases implying deficiencies of cumulus convection. The biases in precipitation-related processes also have fast impact on the large-scale state variables as reported in Xie et al. (2012).

To demonstrate possible processes for the increasing drying biases in the precipitation with the hindcast lead time in the targeted region, we plotted in Fig. 15 the vertical profiles of MSE and  $\text{MSE}^*$  mean biases, as well as bias contribution from each term in Eq. (4) (temperature, moisture, and geopotential) for CAM5 simulations. The biases were computed in reference to ECMWF analysis, and we also plotted the mean profiles of ECMWF analysis in Fig. 15. For simplicity, we converted the unit

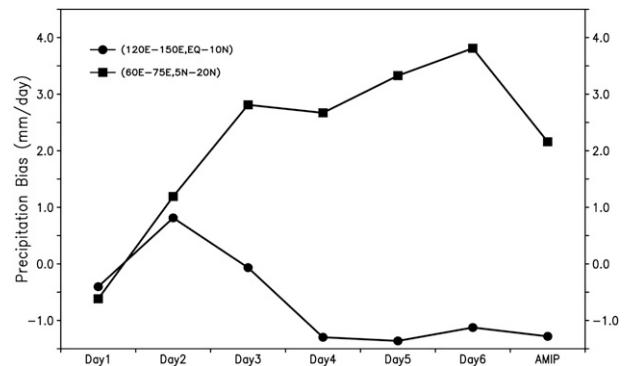
FIG. 12. As in Fig. 11, but for specific humidity bias ( $\text{g kg}^{-1}$ ).

of MSE to temperature (K) by multiplying by  $c_p^{-1}$ . For all the simulations, MSE generally shows positive biases between 500 and 250 hPa and also below 950 hPa, while it shows negative biases between 900 and 600 hPa and also between 200 and 100 hPa. We can further find in the hindcasts that the MSE decreases below 600 hPa and increases between 600 and 300 hPa from day 2 to day 5 as the hindcast lead time increases. For individual contribution of the MSE bias, contribution from geopotential height is very small, and there is no significant tendency in the hindcasts. Both moisture and temperature fields, however, contribute to the total biases, especially for the moisture field. Furthermore, tendencies in the moisture fields contribute mostly to the tendencies in the MSE below 600 hPa and between 600–300 hPa. The biases in  $\text{MSE}^*$  are similar to those shown in MSE but with smaller negative (900–600 hPa) and positive biases (500–250 hPa) above 950 hPa and larger positive biases near the surface. This suggests a larger deficit between the MSE and  $\text{MSE}^*$  profiles in the model compared to the analysis data, especially near the surface. The larger deficit further suggests that the model atmosphere is less saturated and more stable (less CAPE), and is consistent with the negative bias in precipitation over this region.

### b. Wet bias over the Indian Ocean

We now turn our focus to the wet bias over the Indian Ocean. The analyzed region of this wet bias is between

60° and 75°E and between 5° and 10°N. In Fig. 13, CAM5 precipitation for this region generally increases from day 1 to day 6 in the hindcasts. The ECMWF June–August mean  $Q_1$  and  $Q_2$  profiles and associated simulation biases averaged over the selected region are presented in Fig. 16. The model  $Q_1$  profiles generally show warm biases and a warming tendency between 850 and 200 hPa with the hindcast lead time. The model  $Q_2$  profiles show strong drying and a drying tendency between 900 and 700 hPa. These  $Q_1$  and  $Q_2$  warming and drying tendency between day-1 and day-5 hindcasts are consistent with the precipitation bias tendency in Fig. 13.

FIG. 13. CAM5 June–August precipitation mean bias averaged over  $(0^\circ\text{--}10^\circ\text{N}, 120^\circ\text{--}150^\circ\text{E})$  and  $(5^\circ\text{--}20^\circ\text{N}, 60^\circ\text{--}75^\circ\text{E})$  (boxes indicated in Figs. 1, 4).

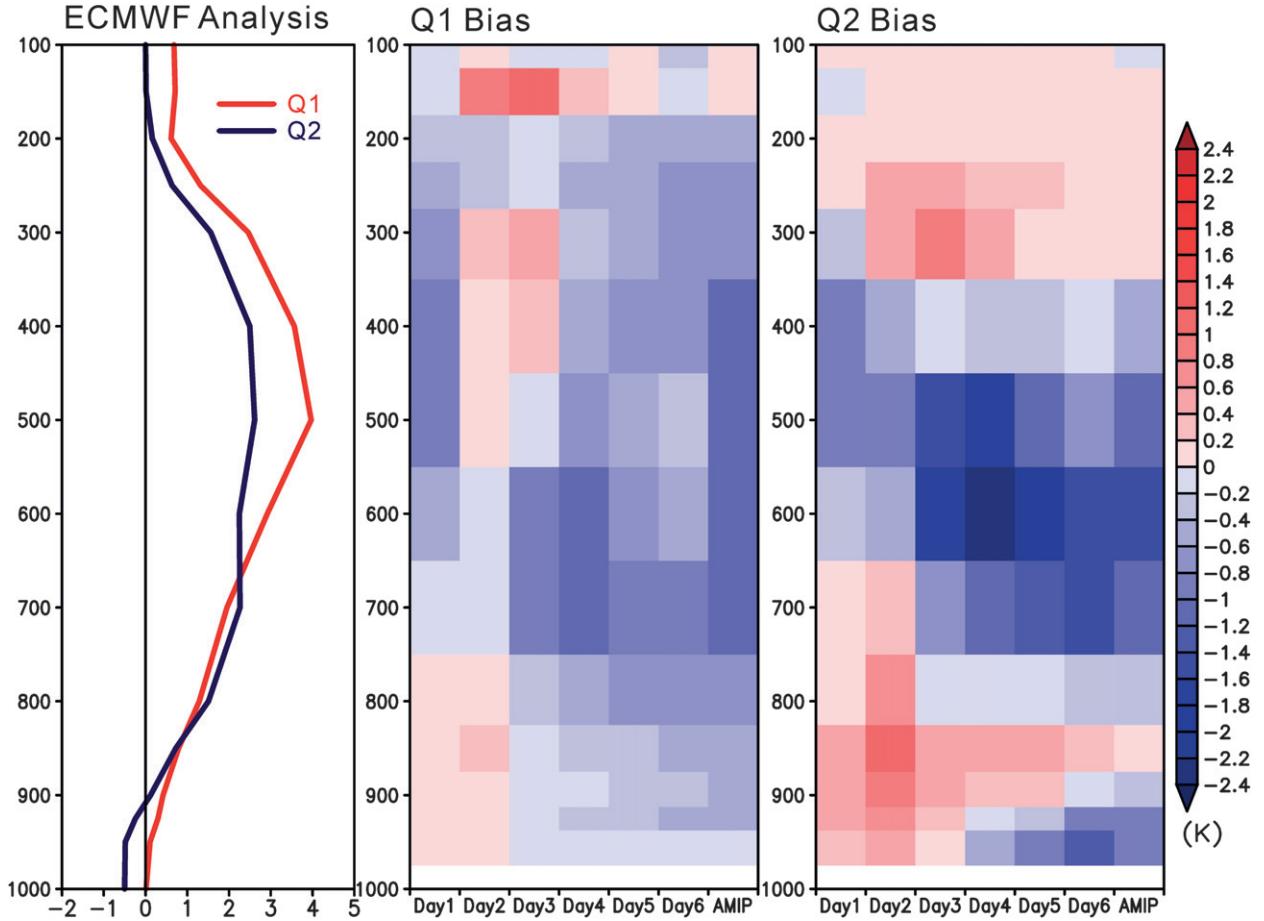


FIG. 14. (left) Composites of profiles of June–August mean  $Q_1$  and  $Q_2$  from ECMWF analysis for the region over  $0^{\circ}$ – $10^{\circ}$ N,  $120^{\circ}$ – $150^{\circ}$ E (boxes indicated in Figs. 1, 4). Also shown here are the CAM5 simulated (middle)  $Q_1$  and (right)  $Q_2$  biases in reference to ECMWF analysis.

Figure 17 shows the vertical profiles of MSE and  $MSE^*$  mean bias, as well as bias contributions from each term for CAM5 simulations. For all the simulations, MSE profiles generally show positive biases below 250 hPa and negative biases above. Between 800 and 600 hPa, slight negative biases are present for the first few days in the hindcast ensembles and follows by increasing in positive biases. Individual contribution of MSE bias indicates that moisture is again the dominant term for MSE bias below 250 hPa. The temperature term shows slightly cold biases for the entire levels, except warm biases are present near the surface and between 500 and 200 hPa in the later hindcasts and AMIP runs. The  $MSE^*$  biases have similar vertical structure to temperature biases, except for larger bias values. From day-1 to day-6 hindcasts, both positive and negative biases in the MSE and  $MSE^*$ , respectively, increase with the hindcast lead time between 800 and 400 hPa. Near the surface, the positive bias is less in  $MSE^*$  than MSE. This suggests

that the model atmosphere is closer to its saturated state and less stable (more CAPE) compared to the analysis, which is consistent with the overestimation of precipitation over this region.

## 5. Summary and discussion

In this study, we proposed and implemented several metrics and diagnostics to systematically examine climate model errors in short-term hindcasts and to quantify how fast hindcast biases approach to climate errors with the emphasis on tropical precipitation and associated moist processes. Our analyses were based on a series of 6-day-long hindcasts and an ensemble of AMIP-type climate simulations (three members) with CAM4 and CAM5 during the YOTC period (May 2008–April 2010). The hindcasts were initialized with ECMWF operational analysis every day at 0000 UTC and prescribed with weekly observed SSTs.

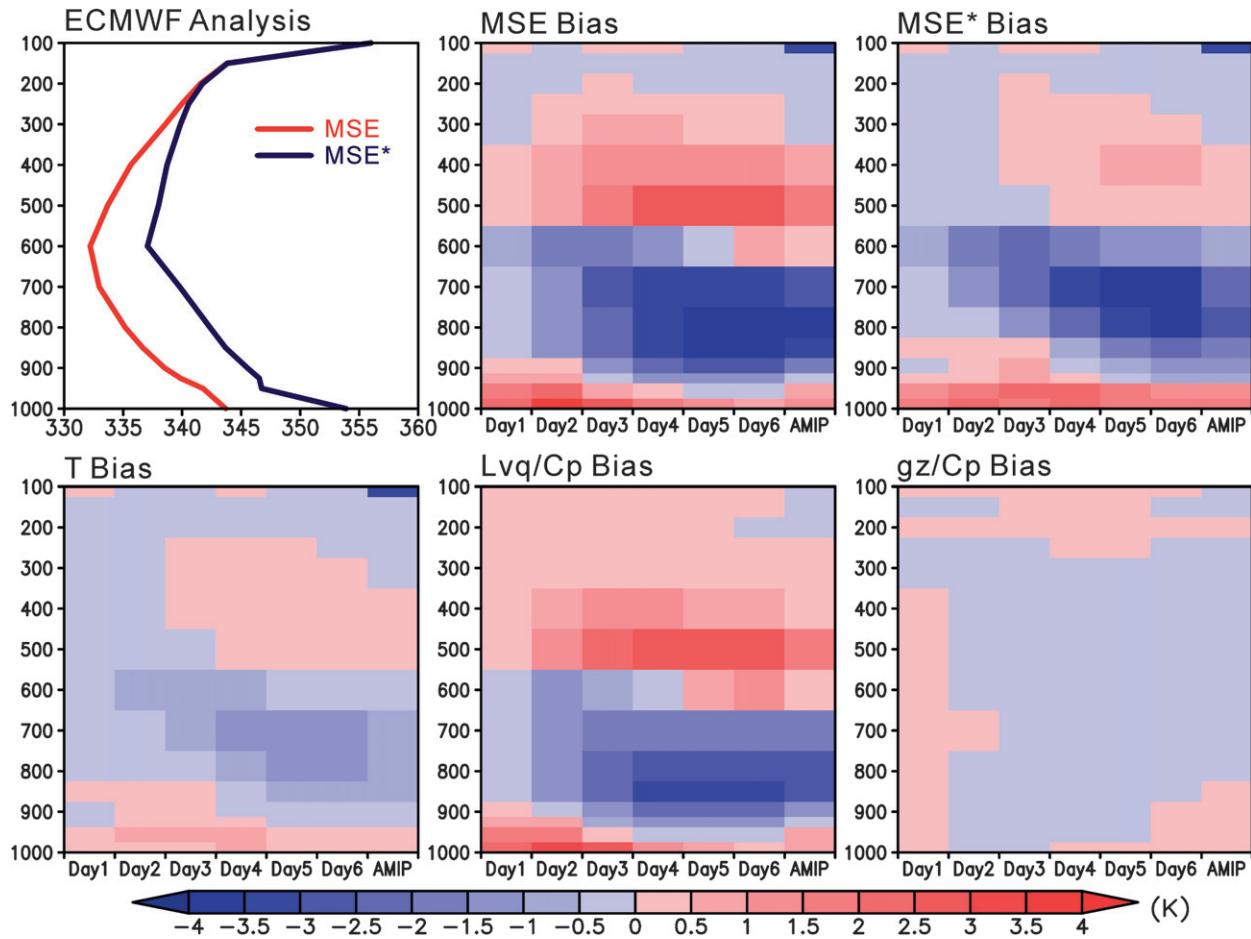
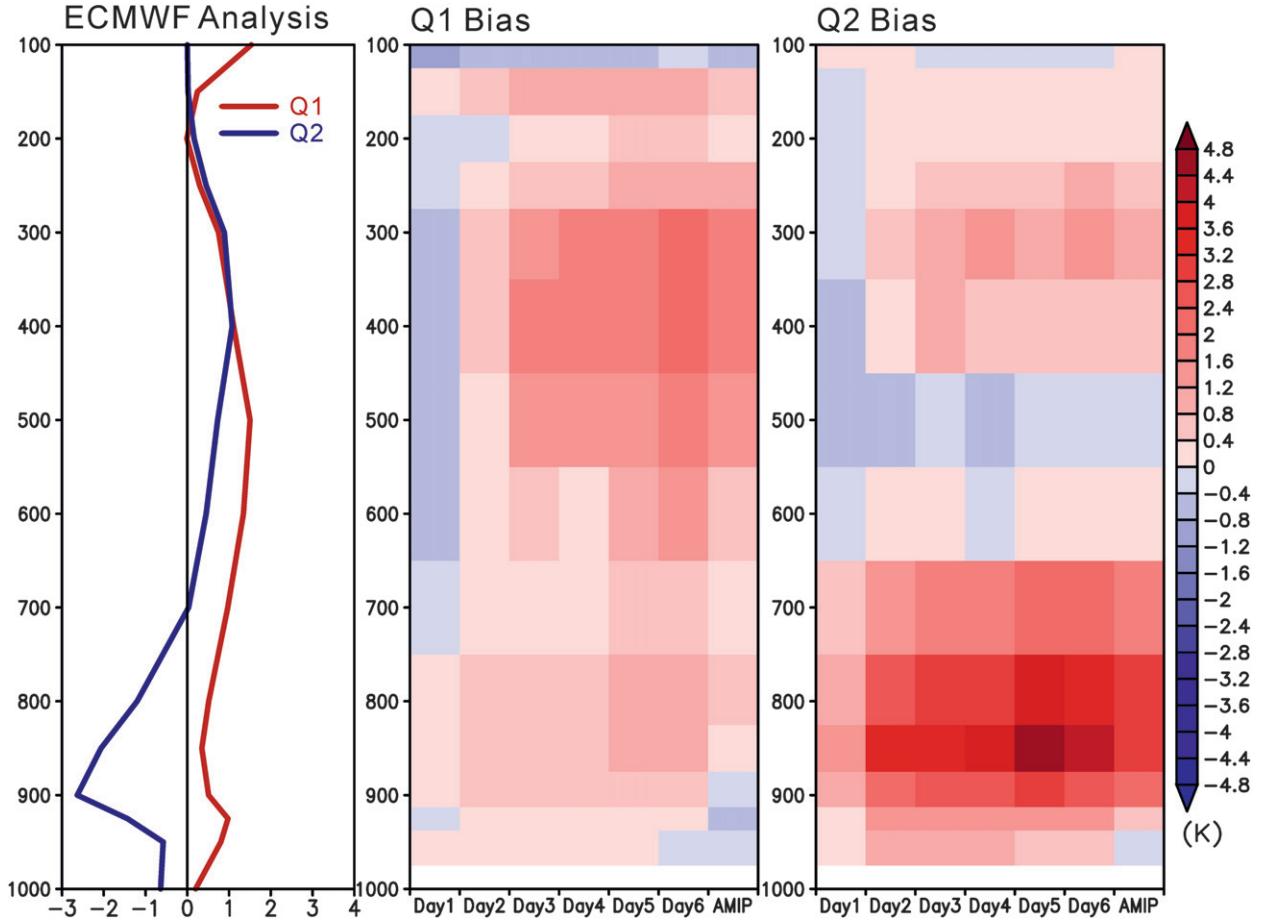


FIG. 15. (top left) June–August mean profiles of ECMWF moist static energy and saturated moist static energy average over  $0^{\circ}$ – $10^{\circ}$ N,  $120^{\circ}$ – $150^{\circ}$ E (boxes indicated in Figs. 1, 4). Also shown here are (top middle), (top right) the biases of these two variables, as well as (bottom left)–(bottom right) the biases from temperature, moisture, and geopotential height from CAM5 hindcasts and AMIP simulations in reference to the ECMWF analysis.

For tropical precipitation evaluation, the performance metrics include precipitation mean bias, RMS errors, pattern correlations, spatial standard deviations, and bias correspondence. We also compiled the following diagnostics: stratiform fraction of precipitation; PDF of daily precipitation intensity; composites of CWV, CRH, temperature, and specific humidity profiles based on daily precipitation intensity; and composites of stratiform rainfall fraction based on CRH. Our results indicate that there is a strong correspondence between short-term hindcasts and long-term climate simulations in precipitation mean biases, bias patterns, fraction of tropical stratiform rainfall, PDF of precipitation intensity, CWV, CRH, and profiles of temperature and specific humidity. All these fields show similar bias patterns or vertical structures in the early hindcasts (fast processes) to those in the long-term AMIP runs.

We also uncovered several model problems in the tropical precipitation-related processes based on these metrics and diagnostics. CAM4 and CAM5 tend to underestimate stratiform rainfall fraction, even though the uncertainty in observations is considered. For the PDFs of daily mean precipitation intensity from both hindcasts and climate runs, both models have higher probability of precipitation intensity than TRMM between  $0.25$  to  $25$   $\text{mm day}^{-1}$  and lower probability for extreme precipitation events ( $>25$   $\text{mm day}^{-1}$ ). The higher probability of precipitation intensity ( $0.25$ – $25$   $\text{mm day}^{-1}$ ) is consistent with the too frequently triggered convection by its deep convection scheme in CAM. This possibly leads to lower moisture in the atmospheric column since column moisture is consumed by precipitation. The relationship between CRH and precipitation intensity further suggests that CAM requires much higher CRH than ECMWF/TRMM to produce similar precipitation

FIG. 16. As in Fig. 14, but for the region over  $5^{\circ}$ – $20^{\circ}$ N,  $60^{\circ}$ – $75^{\circ}$ E.

intensity, especially for extreme precipitation events ( $>25 \text{ mm day}^{-1}$ ). With the too frequently triggered deep convective scheme, the model might not be able to produce such extreme precipitation events. Our further evaluations of the interactions between precipitation and large-scale state variables suggest that the deficit in specific humidity profiles for precipitation intensity larger than  $1 \text{ mm day}^{-1}$  is consistent with overly conversion of specific humidity to rainfall indicated by precipitation PDF. The biases in the temperature profiles, however, do not show clear connection to precipitation biases.

We also performed regional analyses on precipitation in CAM5 simulations over the northwestern Pacific warm pool ( $0^{\circ}$ – $10^{\circ}$ N,  $120^{\circ}$ – $150^{\circ}$ E) and over the Indian Ocean west of the southern Indian peninsula ( $5^{\circ}$ – $10^{\circ}$ N,  $60^{\circ}$ – $75^{\circ}$ E). A dry bias in precipitation is present in the former region, while a wet bias is present in the latter. Both regions are associated with deep convection, and the bias patterns and signs in precipitation are all picked up by both hindcast and AMIP runs, except biases in

hindcast runs are smaller. The precipitation biases are consistent with biases in the vertical profiles of  $Q_1$ , and  $Q_2$ . Based on our diagnostic tools of vertical profiles of MSE and MSE\*, the results suggest that the moisture bias is the dominant contributor to the MSE bias. Furthermore, temperature biases in both regions all show similar bias patterns in the vertical profiles. Precipitation dry bias is associated with a dry tendency near the surface, which allows for less saturated and more stable atmosphere (lower CAPE), while precipitation wet bias is associated with a cold tendency in temperature in the middle and lower troposphere, which allows for more saturated and less stable atmosphere (higher CAPE).

We have demonstrated that our metrics and diagnostics are useful for identifying several key issues in CAM simulations. Although we only performed our analyses over the global tropics and two selected regions during June–August of YOTC, similar techniques can be applied to other regions or time periods with different models, such as those short-term hindcast experiments that were conducted with many CMIP5 models in the

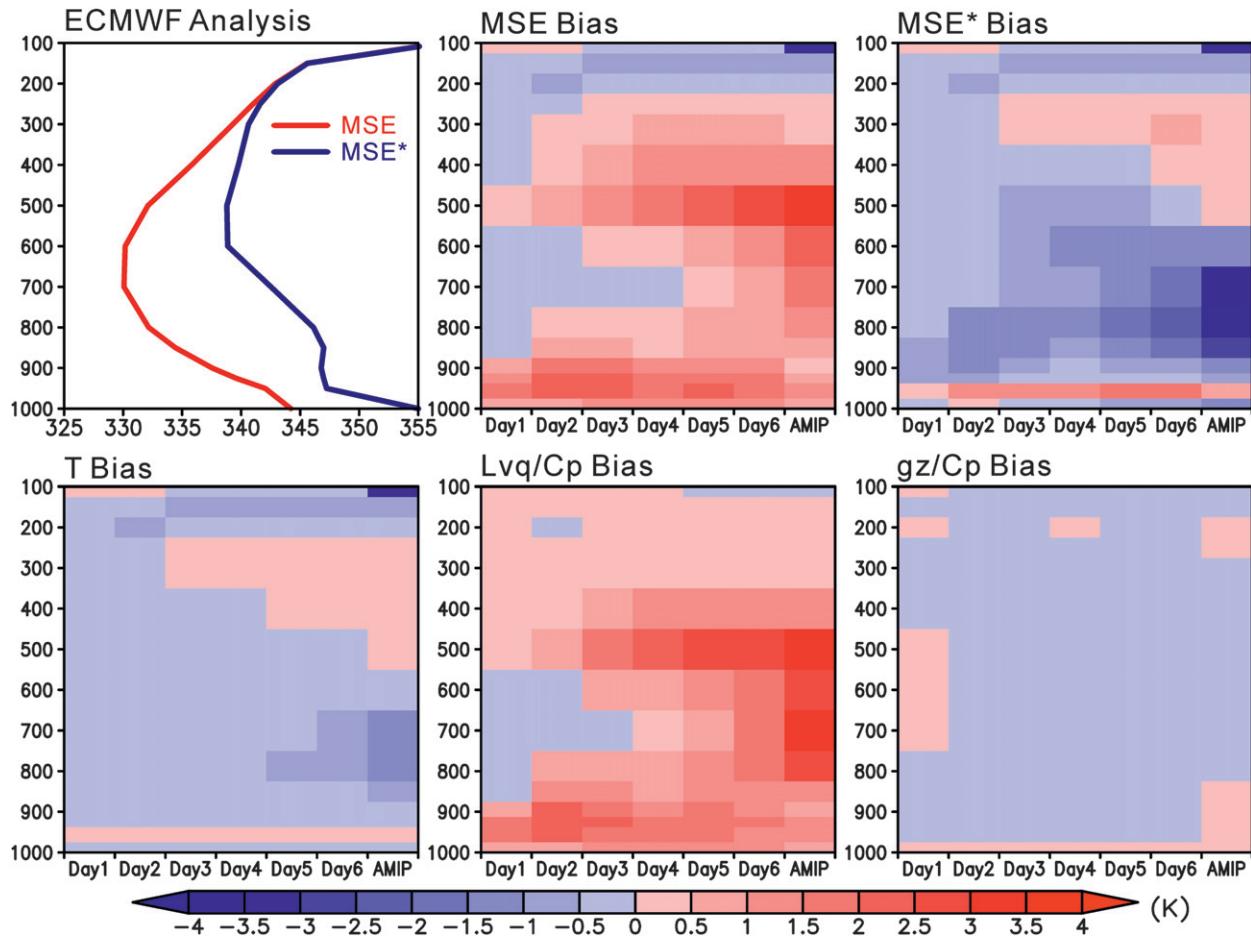


FIG. 17. As in Fig. 15, but for the region over  $0^{\circ}$ – $10^{\circ}$ N,  $120^{\circ}$ – $150^{\circ}$ E.

Transpose-AMIP project. Since climate models are initialized with realistic atmospheric states from NWP analyses in these hindcast experiments, the detailed evolution of parameterized variables in the hindcasts can be compared with detailed field experiment data, and model deficiencies can then be linked directly with specific atmospheric processes observed during field campaigns. Therefore, our future studies will emphasize on developing those process-oriented diagnostics that utilize data from major field programs such as the DOE Atmospheric Radiation Measurement (ARM) Program, as well as applying these metrics and diagnostics to the Transpose-AMIP models, to gain more insights into the cause of climate model systematic errors.

**Acknowledgments.** We are grateful to the ECMWF for making their operational analyses available, and we thank Drs. Yunyan Zhang and Chuanfeng Zhao for collecting the ECMWF-YOTC analysis. We also thank Dr. Yunyan Zhang for very helpful discussion on this paper. Computing resources were provided from the

Livermore Computing Center at Lawrence Livermore National Laboratory (LLNL) and the National Energy Research Scientific Computing Center (NERSC). The efforts of the authors were funded by the Regional and Global Climate Modeling and Atmospheric System Research programs of the U.S. Department of Energy as part of the CAPT. This work was performed under the auspices of the U.S. Department of Energy by LLNL under Contract DE-AC52-07NA27344.

## 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.
- Benedict, J. J., and D. A. Randall, 2007: Observed characteristics of the MJO relative to maximum rainfall. *J. Atmos. Sci.*, **64**, 2332–2354.
- Boo, K.-O., G. Martin, A. Sellar, C. Senior, and Y.-H. Byun, 2011: Evaluating the East Asian monsoon simulation in climate models. *J. Geophys. Res.*, **116**, D01109, doi:10.1029/2010JD014737.
- Boyle, J., and S. A. Klein, 2010: Impact of horizontal resolution on climate model forecasts of tropical precipitation and diabatic

- heating for the TWP-ICE period. *J. Geophys. Res.*, **115**, D23113, doi:10.1029/2010JD014262.
- , and Coauthors, 2005: Diagnosis of Community Atmospheric Model 2 (CAM2) in numerical weather forecast configuration at Atmospheric Radiation Measurement sites. *J. Geophys. Res.*, **110**, D15S15, doi:10.1029/2004JD005042.
- , S. A. Klein, G. Zhang, S. Xie, and X. Wei, 2008: Climate model forecast experiment for TOGA COARE. *Mon. Wea. Rev.*, **136**, 808–832.
- Bretherton, C. S., and S. Park, 2009: A new moist turbulence parameterization in the Community Atmosphere Model. *J. Climate*, **22**, 3422–3448.
- , M. E. Peters, and L. E. Back, 2004: Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, **17**, 1517–1528.
- Chou, C., J. D. Neelin, C.-A. Chen, and J.-Y. Tu, 2009: Evaluating the “rich-get-richer” mechanism in tropical precipitation change under global warming. *J. Climate*, **22**, 1982–2005.
- Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630.
- Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, doi:10.1029/2007JD008972.
- 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.
- Hannay, C., D. L. Williamson, J. J. Hack, J. T. Kiehl, S. A. Klein, C. S. Bretherton, and M. Kohler, 2009: Evaluation of forecasted southeast Pacific stratocumulus in the NCAR, GFDL, and ECMWF models. *J. Climate*, **22**, 2871–2889.
- Holloway, C. E., and J. D. Neelin, 2010: Temporal relations of column water vapor and tropical precipitation. *J. Atmos. Sci.*, **67**, 1091–1105.
- Hurrell, J., G. A. Meehl, D. Bader, T. L. Delworth, B. Kirtman, and B. Wielicki, 2010: A unified modeling approach to climate system prediction. *Bull. Amer. Meteor. Soc.*, **91**, 1819–1832.
- Iacono, M., E. Mlawer, S. Clough, and J.-J. Morcrette, 2000: Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3. *J. Geophys. Res.*, **105**, 14 873–14 890.
- Kim, D., and Coauthors, 2009: Application of MJO simulation diagnostics to climate models. *J. Climate*, **22**, 6413–6436.
- , A. H. Sobel, E. D. Maloney, D. M. W. Frierson, and I.-S. Kang, 2011: A systematic relationship between intraseasonal variability and mean state bias in AGCM simulations. *J. Climate*, **24**, 5506–5520.
- Klein, S. A., X. Jiang, J. Boyle, S. Malyshev, and S. Xie, 2006: Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophys. Res. Lett.*, **33**, L18805, doi:10.1029/2006GL027567.
- Krishnamurti, T. N., and Coauthors, 2003: Improved skill for the anomaly correlation of geopotential heights at 500 hPa. *Mon. Wea. Rev.*, **131**, 1082–1102.
- Lin, J.-L., 2007: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean–atmosphere feedback analysis. *J. Climate*, **20**, 4497–4525.
- , B. Mapes, M. Zhang, and M. Newman, 2004: Stratiform precipitation, vertical heating profiles, and the Madden-Julian oscillation. *J. Atmos. Sci.*, **61**, 296–309.
- Lin, Y., and Coauthors, 2012: TWP-ICE global atmospheric model intercomparison: Convection responsiveness and resolution impact. *J. Geophys. Res.*, **117**, D09111, doi:10.1029/2011JD017018.
- Lintner, B. R., and J. D. Neelin, 2010: Tropical South America/Atlantic sector convective margins and their relationship to low-level inflow. *J. Climate*, **23**, 2671–2685.
- Luo, H., and M. Yanai, 1984: The large-scale circulation and heat sources over the Tibetan Plateau and surrounding areas during the early summer of 1979. Part II: Heat and moisture budgets. *Mon. Wea. Rev.*, **112**, 130–141.
- Mapes, B. E., R. Milliff, and J. Morzel, 2009: Composite life cycle of maritime tropical mesoscale convective systems in scatterometer and microwave satellite observations. *J. Atmos. Sci.*, **66**, 199–208.
- Martin, G. M., S. F. Milton, C. A. Senior, M. E. Brooks, S. Ineson, T. Reichler, and J. Kim, 2010: Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate. *J. Climate*, **23**, 5933–5957.
- Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682.
- Morita, J., Y. N. Takayabu, S. Shige, and Y. Kodama, 2006: Analysis of rainfall characteristics of the Madden-Julian oscillation using TRMM satellite data. *Dyn. Atmos. Oceans*, **42**, 107–126.
- Morrison, H., and A. Gettelman, 2008: A new two-moment bulk stratiform cloud microphysics scheme in the NCAR Community Atmosphere Model (CAM3). Part I: Description and numerical tests. *J. Climate*, **21**, 3642–3659.
- Neale, R. B., J. H. Richter, and M. Jochum, 2008: The impact of convection on ENSO: From a delayed oscillator to a series of events. *J. Climate*, **21**, 5904–5924.
- , —, S. Park, P. H. Lauritzen, S. J. Vavrus, P. J. Rasch, and M. Zhang, 2013: The mean climate of the Community Atmosphere Model (CAM4) in forced SST and fully coupled experiments. *J. Climate*, in press.
- Neelin, J. D., O. Peters, and K. Hales, 2009: The transition to strong convection. *J. Atmos. Sci.*, **66**, 2367–2384.
- Palmer, T. N., F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell, 2008: Toward seamless prediction: Calibration of climate change projections using seasonal forecasts. *Bull. Amer. Meteor. Soc.*, **89**, 459–470.
- Park, S., and C. S. Bretherton, 2009: The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model. *J. Climate*, **22**, 3449–3469.
- Phillips, T. J., and Coauthors, 2004: Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bull. Amer. Meteor. Soc.*, **85**, 1903–1915.
- Pincus, R., C. P. Batstone, R. J. P. Hofmann, K. E. Taylor, and P. J. Gleckler, 2008: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *J. Geophys. Res.*, **113**, D14209, doi:10.1029/2007JD009334.
- Reichler, T., and J. Kim, 2008: How well do coupled models simulate today’s climate? *Bull. Amer. Meteor. Soc.*, **89**, 303–311.
- 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.

- Rodwell, M. J., and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Quart. J. Roy. Meteor. Soc.*, **133**, 129–146, doi:10.1002/qj.23.
- Schumacher, C., and R. A. Houze Jr., 2003: Stratiform rain in the tropics as seen by the TRMM precipitation radar. *J. Climate*, **16**, 1739–1756.
- Seo, K.-H., J.-K. E. Schemm, C. Jones, and S. Moorthi, 2005: Forecast skill of the tropical intraseasonal oscillation in the NCEP GFS dynamical extended range forecasts. *Climate Dyn.*, **25**, 265–284, doi:10.1007/s00382-005-0035-2.
- Sherwood, S. C., R. Roca, T. M. Weckwerth, and N. G. Andronova, 2010: Tropospheric water vapor, convection, and climate. *Rev. Geophys.*, **48**, RG2001, doi:10.1029/2009RG000301.
- Sperber, K. R., and D. Kim, 2012: Simplified metrics for the identification of the Madden–Julian oscillation in models. *Atmos. Sci. Lett.*, **13**, 187–193.
- Stephens, G. L., and Coauthors, 2002: The CloudSat mission and the A-Train. *Bull. Amer. Meteor. Soc.*, **83**, 1771–1790.
- Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192.
- , R. J. Stouffer, G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498.
- Waliser, D. E., and Coauthors, 2012: The “Year” of Tropical Convection (May 2008–April 2010): Climate variability and weather highlights. *Bull. Amer. Meteor. Soc.*, **93**, 1189–1218.
- Wang, B., H.-J. Kim, K. Kikuchi, and A. Kitoh, 2011: Diagnostic metrics for evaluation of annual and diurnal cycles. *Climate Dyn.*, **37**, 941–955, doi:10.1007/s00382-010-0988-0.
- Williams, K. D., and M. E. Brooks, 2008: Initial tendencies of cloud regimes in the Met Office Unified Model. *J. Climate*, **21**, 833–840.
- , and Coauthors, 2013: The Transpose-AMIP II experiments and its application to the understanding of Southern Ocean cloud biases in climate models. *J. Climate*, in press.
- Williamson, D. L., and J. G. Olson, 2007: A comparison of forecast errors in CAM2 and CAM3 at the ARM Southern Great Plains site. *J. Climate*, **20**, 4572–4585.
- , and Coauthors, 2005: Moisture and temperature balances at the Atmospheric Radiation Measurement Southern Great Plains Site in forecasts with the Community Atmosphere Model (CAM2). *J. Geophys. Res.*, **110**, D15S16, doi:10.1029/2004JD005109.
- Xie, S., and M. Zhang, 2000: Impact of the convective triggering function on single-column model simulations. *J. Geophys. Res.*, **105**, 14 983–14 996.
- , —, J. S. Boyle, R. T. Cederwall, G. L. Potter, and W. Lin, 2004: Impact of a revised convective triggering mechanism on Community Atmosphere Model, Version 2, simulations: Results from short-range weather forecasts. *J. Geophys. Res.*, **109**, D14102, doi:10.1029/2004JD004692.
- , J. Boyle, S. A. Klein, X. Liu, and S. Ghan, 2008: Simulations of Arctic mixed-phase clouds in forecasts with CAM3 and AM2 for M-PACE. *J. Geophys. Res.*, **113**, D04211, doi:10.1029/2007JD009225.
- , H.-Y. Ma, J. S. Boyle, S. A. Klein, and Y. Zhang, 2012: On the correspondence between short- and long-time-scale systematic errors in CAM4/CAM5 for the Year of Tropical Convection. *J. Climate*, **25**, 7937–7955.
- Yanai, M., and T. Tomita, 1998: Seasonal and interannual variability of atmospheric heat sources and moisture sinks as determined from NCEP–NCAR reanalysis. *J. Climate*, **11**, 463–482.
- , S. Esbensen, and J.-H. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30**, 611–627.
- Yang, F., 2011: Review of NCEP GFS forecast skills and major upgrades. Preprints, 24th Conf. on Weather and Forecasting/20th Conf. on Numerical Weather Prediction, Seattle, WA, Amer. Meteor. Soc., 2B.1. [Available online at <https://ams.confex.com/ams/91Annual/flvgateway.cgi/id/17618?recordingid=17618>.]
- Yokoi, S., and Coauthors, 2011: Application of cluster analysis to climate model performance metrics. *J. Appl. Meteor. Climatol.*, **50**, 1666–1675.
- Zhang, G. J., and N. A. McFarlane, 1995: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Center general circulation model. *Atmos.–Ocean*, **33**, 407–446.