⊘Using the Atmospheric Radiation Measurement (ARM) Datasets to Evaluate Climate Models in Simulating Diurnal and Seasonal Variations of Tropical Clouds⊘

HAILONG WANG, CASEY D. BURLEYSON, PO-LUN MA, JEROME D. FAST, AND PHILIP J. RASCH

Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington

ABSTRACT

Long-term Atmospheric Radiation Measurement (ARM) datasets collected at the three tropical western Pacific (TWP) sites are used to evaluate the ability of the Community Atmosphere Model (CAM5) to simulate the various types of clouds, their seasonal and diurnal variations, and their impact on surface radiation. A number of CAM5 simulations are conducted at various horizontal grid spacing (around 2°, 1°, 0.5°, and 0.25°) with meteorological constraints from analysis or reanalysis. Model biases in the seasonal cycle of cloudiness are found to be weakly dependent on model resolution. Positive biases (up to 20%) in the annual mean total cloud fraction appear mostly in stratiform ice clouds. Higher-resolution simulations do reduce the positive bias in ice clouds, but they inadvertently increase the negative biases in convective clouds and low-level liquid clouds, leading to a positive bias in annual mean shortwave fluxes at the sites, as high as $65\,\mathrm{W\,m^{-2}}$ in the 0.25° simulation. Such resolution-dependent biases in clouds can adversely lead to biases in ambient thermodynamic properties and, in turn, produce feedback onto clouds. Both the model and observations show distinct diurnal cycles in total, stratiform, and convective cloud fractions; however, they are out of phase by 12 h and the biases vary by site. The results suggest that biases in deep convection affect the vertical distribution and diurnal cycle of stratiform clouds through the transport of vapor and/or the detrainment of liquid and ice. The approach used here can be easily adapted for the evaluation of new parameterizations being developed for CAM5 or other global or regional models.

1. Introduction

Atmospheric moist convection in the tropics, which is intimately tied to large-scale circulations, redistributes heat, moisture, and momentum globally. Marine boundary layer convection and tropical clouds, which are key elements of the global energy balance and water cycle, remain an important source of uncertainty in our understanding of tropical cloud feedback processes, climate sensitivity, and even predictions of changes in the climate system (e.g., Bony and Dufresne 2005). It is a significant

challenge for weather and climate models that use cumulus parameterizations to represent the microphysical, thermodynamical, and dynamical processes of convective clouds. Many current models still have difficulty representing shallow convection and the onset of deep convection, which limits our ability to understand and predict changes in the water cycle in a warmer climate. The tropical western Pacific (TWP) warm pool is one region that plays an important role in the global climate system. This area is characterized by high sea surface temperatures (SSTs), humid air, strong solar heating, and active convection with clouds and precipitation that vary on a range of time scales (e.g., diurnal, intraseasonal, seasonal, and interannual variations). Convection and cloud variability associated with the Madden–Julian oscillation (MJO; Madden and Julian 1994) and El Niño–Southern Oscillation (ENSO) have a broader and more profound impact on the energy budget and hydrologic cycle in the global climate system.

Climate models often have large biases in predicting clouds and their radiative effects over the TWP

(e.g., Qian et al. 2012) as well as tropical clouds in general (e.g., Wang and Su 2013). More specifically, the recent generations of climate models had been shown to underestimate the fractional coverage of tropical low clouds but to overestimate their optical depth and radiative effects (e.g., Zhang et al. 2005; Nam et al. 2012), which is commonly referred to as the "too few, too bright" tropical low-cloud problem. Compensating errors in the estimates of cloud macrophysical and optical properties may reduce biases in the radiation budget but may also hide problems in model parameterizations that can impede our ability to understand climate responses to anthropogenic perturbations (e.g., Bony and Dufresne 2005; Medeiros et al. 2008). Long-term observations are critical for understanding the root causes of these errors and/or biases. For most of the past two decades, the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Program collected data at three surface sites in the TWP—Manus (Papua New Guinea), Nauru (Republic of Nauru), and Darwin (Australia)—with the goal of creating a climate data record in a sparsely sampled region (Mather et al. 1998; Ackerman et al. 1999; Long et al. 2013). The ARM TWP sites cover a large geographical area in the trade-wind region and are likely representative of the broader tropics (Jakob et al. 2005). Long-term ARM measurements collected over the TWP have been used to understand the linkages among the ocean, tropical convection, cloud properties, and radiation as well as to characterize the vertical transport of moisture, heat, and momentum in convection (Long et al. 2013). The ARM datasets provide an excellent resource for evaluating climate models using both statistical and process-oriented approaches and for eventually reducing model errors in cloud parameterizations. For example, McFarlane et al. (2013) analyzed the datasets to quantify the surface radiation budget and cloud radiative effects by cloud types and study their intraseasonal and interannual variability. Burleyson et al. (2015) further analyzed the diurnal cycles of cloud populations and their radiative effects.

In the present study, we establish a procedure for using the ARM TWP datasets to evaluate the ability of the Community Atmosphere Model, version 5 (CAM5), to simulate the various cloud types and their seasonal and diurnal variations, as well as meteorological fields and cloud radiative effects. The long-term multisensor ARM datasets allow us to document and investigate model biases in a more complete manner than previous studies. For example, we can use the sounding datasets to document biases in the model temperature and moisture profiles, which in turn may lead to errors in simulating specific types of clouds and biases in the surface radiation budget. The analyses in this paper demonstrate how we

can best use the long-term ARM datasets to identify the dominant sources of model biases and uncertainties. This approach can be easily adapted for the evaluation of other global or regional models.

Similar approaches have been used in previous studies to evaluate model climatology and/or forecast of clouds and the surface radiation budget over other long-term observing sites (e.g., Hinkelman et al. 1999; Hogan et al. 2001; Illingworth et al. 2007; Paquin-Ricard et al. 2010). Qian et al. (2012) showed that the free-running version of CAM5, like many other climate models, captures the climatological seasonal variations in total cloudiness but has a significant high bias in total cloud fraction and a cold temperature bias in the upper troposphere over the TWP Manus site. Our CAM5 simulations are designed to minimize the impact of some known biases and potential circulation errors. Earlier studies (e.g., Comstock and Jakob 2004; Boyle et al. 2005) have also shown that global models have difficulties in simulating the vertical distribution of clouds over the TWP sites. Chandra et al. (2015) found that the CAM5 running in 6-day hindcast mode underestimates low clouds (those forming below the freezing level) in general, but overproduces low clouds in a thin layer near the top of the mixed layer over the ARM Manus site as a result of excessive humidity. Compared to the previous studies, this present study uses a different modeling and analysis approach and adds new insights into model deficiencies and model–observation comparison. To minimize the impact of the known CAM5 biases in meteorology on clouds, the modeled meteorological fields in some of our simulations are nudged toward an atmospheric analysis or reanalysis dataset to various degrees and evaluated against sounding measurements. The details of these methods are summarized in the following methodology section. Simulations are conducted at various resolutions to examine the resolution dependency of modeled cloud statistics. Modeled and observed clouds are further decomposed into different types (e.g., convective vs stratiform and liquid vs ice) to provide insights into the performance of specific cloud parameterization schemes for the TWP region.

2. Methodology

a. ARM TWP measurements

The ARM TWP datasets used to evaluate the climate model in this study were developed by McFarlane et al. (2013) and Burleyson et al. (2015). Monthly averaged values are used to evaluate the seasonal cycle simulated by the model. For the evaluation of the diurnal cycle on the 2°-grid model, data are averaged from their native

TABLE 1. ARM measurements and data streams at the TWP sites. The DOI assigned to each ARM VAP is included for reference.

Site (lat, lon)	Measurement	ARM VAP data and source	Earliest date	Latest date	Native temporal resolution	No. of 3-h mean samples
Manus (2.06°S, 147.43°E)	Cloud type and fraction profiles	ARSCL https://doi.org/10.5439/1027282	1 Jan 2002	2 Dec 2013	2 min	28 744
	Liquid water path	ARMBECLDRAD https://doi.org/10.5439/1039930	3 Mar 1999	9 Apr 2010	1 h	26 660
	Meteorological profiles	SONDEWNPN https://doi.org/10.5439/1021460	18 Aug 2001	7 Jul 2014	Intermittent	10 168
	Surface radiation	QCRAD https://doi.org/10.5439/1027372	31 Oct 1998	26 Apr 2014	1 min	44 300
Nauru (0.52°S, 166.92°E)	Cloud type and fraction profiles	ARSCL https://doi.org/10.5439/1027282	1 Jan 2002	14 Feb 2009	2 min	14 776
	Liquid water path	ARMBECLDRAD https://doi.org/10.5439/1039928	3 Mar 1999	9 Apr 2010	1 h	28 424
	Meteorological profiles	SONDEWNPN https://doi.org/10.5439/1021460	5 Nov 1998	25 Aug 2013	Intermittent	10 564
	Surface radiation	QCRAD https://doi.org/10.5439/1027372	31 Oct 1998	31 Aug 2013	1 min	43 161
Darwin (12.42°S, 130.89°E)	Cloud type and fraction profiles	ARSCL https://doi.org/10.5439/1027282	2 Jan 2006	31 Dec 2013	2 min	16 545
	Liquid water path	ARMBECLDRAD https://doi.org/10.5439/1039929	1 Apr 2002	9 Apr 2010	1 h	15 164
	Meteorological profiles	SONDEWNPN https://doi.org/10.5439/1021460	1 Apr 2002	25 Jul 2014	Intermittent	17 886
	Surface radiation	QCRAD https://doi.org/10.5439/1027372	12 Mar 2002	20 Apr 2014	1 min	35 357

temporal resolution into 3-h time windows to better compare with statistics within one 2° grid box or column of the climate model. Assuming wind speeds on the order of 15 m s^{-1}, a typical air mass in this trade-wind region would cover approximately 160 km in a 3-h window, roughly equivalent to the 2° grid box or column. Across all time scales we examine, the long data record at the TWP sites ensures that adequate sample sizes are available for every dataset that we used to evaluate the model's cloud and radiation climatology (Table 1). We use four primary data streams in our analysis, the details of which are provided in Table 1.

The profiles of cloud fraction and cloud type are based on the Active Remote Sensing of Clouds (ARSCL; Clothiaux et al. 2001) value-added product (VAP). Cloud type classifications are done using the simple scheme based on cloud bases and tops described in McFarlane et al. (2013; their Table 3) and utilized in Burleyson et al. (2015; their Table 2). Our processing of this data stream is the same as in those two studies, including the exclusion of all vertical profiles in which rain rate at the cloud base is larger than 1 mm h^{-1} to avoid data quality issues during periods of likely attenuation of the active sensors. Using the categories of McFarlane et al. (2013), we define convective clouds in the observational datasets as "low" (mostly cumulus), cumulus congestus, and deep convection. Stratiform clouds include altocumulus (not identified

as convective because they are not rooted in the boundary layer), altostratus, cirrostratus, and cirrus. The separation of liquid and ice clouds is based on the phase retrieval included in the ARSCL product.

Liquid water path (LWP) estimates are based on microwave radiometer retrievals and were processed into 1-h resolution average values as part of the ARM Best Estimate Cloud Radiation Measurements (ARMBECLDRAD; Xie et al. 2010) VAP. The time series of clear-sky and all-sky downwelling shortwave (SW) and longwave (LW) radiative fluxes at the surface are based on the Data Quality Assessment for ARM Radiation Data (QCRAD) VAP, which is described in Long and Ackerman (2000) and Long and Turner (2008). Measurements of the vertical profiles of temperature, moisture, and wind come from sounding data processed and named the SONDEWNPN data stream. Vertical profiles from the sounding data were averaged into 5-hPa vertical bins and then interpolated to model levels to facilitate comparison with the model output. Note that the temporal sampling of sounding was very limited and thus the data are unable to characterize the diurnal variations in meteorological profiles.

b. Model description and experiments

The global climate model used for testing and evaluation is the CAM5, which is the atmospheric component

TABLE 2. A summary of the CAM5 experiments and model configurations used in our analysis. Simulation type describes whether the model is performed in the specified dynamics (SD), free running (FR), or nudged winds (NG) mode. The two FR simulations were prescribed with present-day climatological mean SSTs.

	Expt	Horizontal grid (lat × lon)	Simulation type	Year(s)
Set 1	2° grid	1.9° × 2.5°	SD	2009
	1° grid	0.9° × 1.25°	SD	2009
	0.5° grid	0.47° × 0.63°	SD	2009
	0.25° grid	0.23° × 0.31°	SD	2009
Set 2	CAM5_2	1.9° × 2.5°	FR	10 yr
	CAM5_1	0.9° × 1.25°	FR	10 yr
	CAM5_2NG	1.9° × 2.5°	NG	2002–11

of the Community Earth System Model (Hurrell et al. 2013). CAM5 has relatively comprehensive representations of clouds and radiation, as well as mechanisms for their interactions with climate (Neale et al. 2012), including a double-moment cloud microphysics scheme (Morrison and Gettelman 2008), a cloud macrophysics scheme for stratiform clouds (Park et al. 2014), a moist turbulence scheme that can explicitly simulate stratus–radiation–turbulence interactions (Bretherton and Park 2009), a shallow convection scheme (Park and Bretherton 2009), longwave and shortwave radiation schemes (Iacono et al. 2008), and a modified Zhang and McFarlane (1995) deep convection scheme that includes subgrid convective momentum transports and an updated closure (Neale et al. 2008; Richter and Rasch 2008). As in many other climate models, stratiform and convective clouds are parameterized separately in CAM5. Four independent cloud fractions are diagnosed within each model grid box, including single-phase shallow and deep cumulus fractions calculated using mass fluxes as well as separate liquid and ice stratus fractions diagnosed using the averaged relative humidity in the noncumulus portion of the grid (Park et al. 2014). When the grid total cloud fraction (i.e., the horizontal areal coverage) is calculated, it is assumed that there is no overlap between shallow and deep cumulus clouds, maximum overlap between liquid and ice stratus clouds, and no overlap between cumulus and stratus clouds. For radiative transfer calculations, total cloud fractions of individual model layers are vertically integrated using a maximum random overlap assumption (Park et al. 2014).

We conducted two sets of CAM5 simulations that are summarized in Table 2. The experiment design of the first set of simulations is documented in Ma et al. (2015) and is briefly described here. This set is conducted to test the resolution sensitivity of clouds and radiation. The model was configured to run at horizontal grid spacings of 1.9° × 2.5°, 0.9° × 1.25°, 0.47° × 0.63°, or 0.23° × 0.31°

(hereafter 2°, 1°, 0.5°, and 0.25° horizontal grids, respectively) and 30 vertical levels with a model time step of 15 min in the SD mode. The model meteorological fields (winds, surface pressure, surface stress, and surface heat and moisture fluxes) are constrained using a methodology described by Ma et al. (2013) to agree with the high-resolution (0.15°) ECMWF Year of Tropical Convection analysis (Dee et al. 2011; Waliser et al. 2012), while atmospheric constituents including water vapor, clouds, and aerosols are allowed to evolve according to the physics in the model. This SD approach has been used to minimize the impact of potential biases in meteorology on clouds. Furthermore, we intend to test the performance of cloud parameterizations at the different model horizontal resolutions by keeping the meteorological conditions as close to reality as possible. These resolution-sensitivity experiments were conducted from 1 November 2008 to 1 January 2010 and one year of model output in 2009 is used in our analyses.

A second set of simulations was designed to model the present-day climatology, focusing on diurnal cycles of clouds and surface radiative fluxes that can be evaluated against the long-term ARM TWP datasets. Two free-running simulations, CAM5_2 and CAM5_1, were conducted for 11 years at 2° and 1° horizontal grid spacing, respectively, using present-day initial and boundary conditions (e.g., the climatological monthly mean sea surface temperatures and sea ice concentrations). We used monthly mean fields from the last 10 years in our analysis. Another 11-yr (2001–11) simulation at 2° horizontal grid spacing, CAM5_2NG, was performed with horizontal winds being nudged to agree with ERA-Interim (Dee et al. 2011) with a 6-h relaxation time scale following previous studies (e.g., Ma et al. 2013; Zhang et al. 2014). The corresponding monthly sea surface temperatures and sea ice concentrations were prescribed. Monthly and 3-hourly mean model fields for 10 years are used for the analysis and evaluation of seasonal and diurnal cycles in the model. Note that in the CAM5_2NG experiment only winds are nudged to reanalysis while the other meteorological fields are not constrained. We decided to use a less strong observational constraint in CAM5_2NG than in the 1-yr SD simulations (i.e., circulation nudging instead of specified dynamics) because the long-term climatological characteristics of clouds and radiative fluxes should be robust enough and it is unnecessary to possibly introduce unintentional impacts on the model basic states and physics through the SD forcing terms (e.g., Lin et al. 2016).

3. Results

In this section, we first use the 1-yr SD simulations (see Table 2) at four different horizontal resolutions

(i.e., 2°, 1°, 0.5°, and 0.25°) to evaluate model simulated seasonality of clouds and its dependence on model resolution at the three TWP sites. Since these simulations were conducted for only one year, we then looked at the seasonal cycles in three 10-yr simulations with different resolutions and/or meteorological nudging method (CAM5_1, CAM5_2, and CAM5_2NG) and summarize the differences between the 10- and the 1-yr simulations. The four 1-yr simulations are also analyzed for the seasonal cycle of vertical cloud distributions and the impact of model resolution, and the total cloud fraction is further decomposed into convective and stratiform (ice and liquid) cloud fraction in the 10-yr nudged simulation (CAM5_2NG) and compared to the long-term observed ones in section 3b. Although the different meteorological nudging methods are employed in the simulations, systematic biases are seen in the simulated cloud fields. Sounding profiles at the three TWP sites are used to evaluate the temperature and moisture profiles in the model simulations, and results are discussed in section 3c. Last, the diurnal variations of clouds and radiation based on high-frequency model fields in CAM5_2NG are evaluated against long-term ARM observations at the three TWP sites. Note that mean values at the closest grid point to the corresponding ARM sites are used for the comparison with site observations.

a. Seasonality of clouds and dependence on model resolution

The 1-yr CAM5 simulations at four different horizontal resolutions with model meteorological fields specified to the ECMWF analysis are used here to evaluate the dependence of cloud seasonality on model resolution. Figure 1 shows the observed and CAM5 simulated seasonal variation of clouds and downwelling SW and LW fluxes at the Manus site in 2009. The annual mean values of these quantities are summarized in Table 3 for all three sites. In this particular year, the 0.25° simulation has the lowest bias (1%) in annual mean total cloud fraction at the Manus site (Table 3), although the model misses the seasonal cycles at all resolutions (i.e., the observed minimum of cloud amount in late spring and the maximum in summer; Fig. 1a). However, the model reproduces the all-sky downwelling SW flux better at coarser resolutions. Larger SW radiative fluxes reaching the surface at finer resolutions are consistent with the smaller LWPs in the model. The 0.25° simulation has the highest bias in the downwelling SW flux and LWP, +65 W m^{-2} and −12 g m^{-2}, respectively. However, opposite results are found for both all-sky (LWdn; Fig. 1e) and clear-sky downwelling LW fluxes at the surface (LWdn_c; Fig. 1f), with the 0.25° simulation

producing the smallest bias. The significant bias in LW fluxes in these constrained simulations is likely due to a mismatch between the point measurements and the model gridbox mean. The grid box containing the Manus site is partially occupied by ocean that normally has warmer temperatures (especially at night) and greater upwelling or downwelling LW fluxes compared to the island. However, this effect cannot explain why the finer-resolution simulations give a smaller bias in LW fluxes. The model captures the seasonal cycle of clear-sky SW radiative fluxes at the surface, which has little relation to model performance but serves as a sanity check on other key components of the model (e.g., aerosols, water vapor, and gaseous absorption).

The sensitivity of clouds and radiation to model resolution at the Nauru and Darwin ARM sites (Figs. S1 and S2 in the supplemental material) is similar to but smaller in magnitude than at Manus. However, it should be noted that cloud measurements are missing for most months in 2009 at these two sites. The model also has an overall smaller bias in surface radiation over Nauru and Darwin than at Manus. The distinct seasonal cycle at the Darwin site is especially well captured by the model (Fig. S2).

Comparing the year 2009 (Fig. 1) to the 10-yr average (Fig. 2), the observations do differ significantly but the model results (for different number of years at 2° resolution) are not statistically different (at the 0.05 significance level), indicating that the model does not adequately capture the observed interannual variability in seasonal cycles. This may result from the fact that 2009 was a moderate El Niño year whereas the free-running simulations were forced with climatological mean SSTs. Figure 2 also shows the contrasts between the two free-running simulations, CAM5_1 and CAM5_2, conducted at a 1° and 2° horizontal grid spacing, respectively. The annual mean values of cloud properties and radiation fluxes are summarized in Table 4 for all three sites. The higher-resolution run (CAM5_1) does have a smaller annual mean bias in LWP and radiative fluxes and captures their seasonality better than the lower-resolution one (CAM5_2) at the Manus site. At the same lower resolution, the nudged simulation (CAM5_2NG) has better performance than the free-running simulation (CAM5_2) does in terms of both annual means and seasonal variations. At the Nauru site (Table 4 and Fig. 3), the modeled annual mean and seasonal variation of cloud and radiation properties have a similar resolution dependence to that at the Manus site. However, it is quite different at the Darwin site (Table 4 and Fig. 4), likely resulting from the distinct contrast in the ambient conditions and clouds associated with monsoon circulations. The meteorological nudging
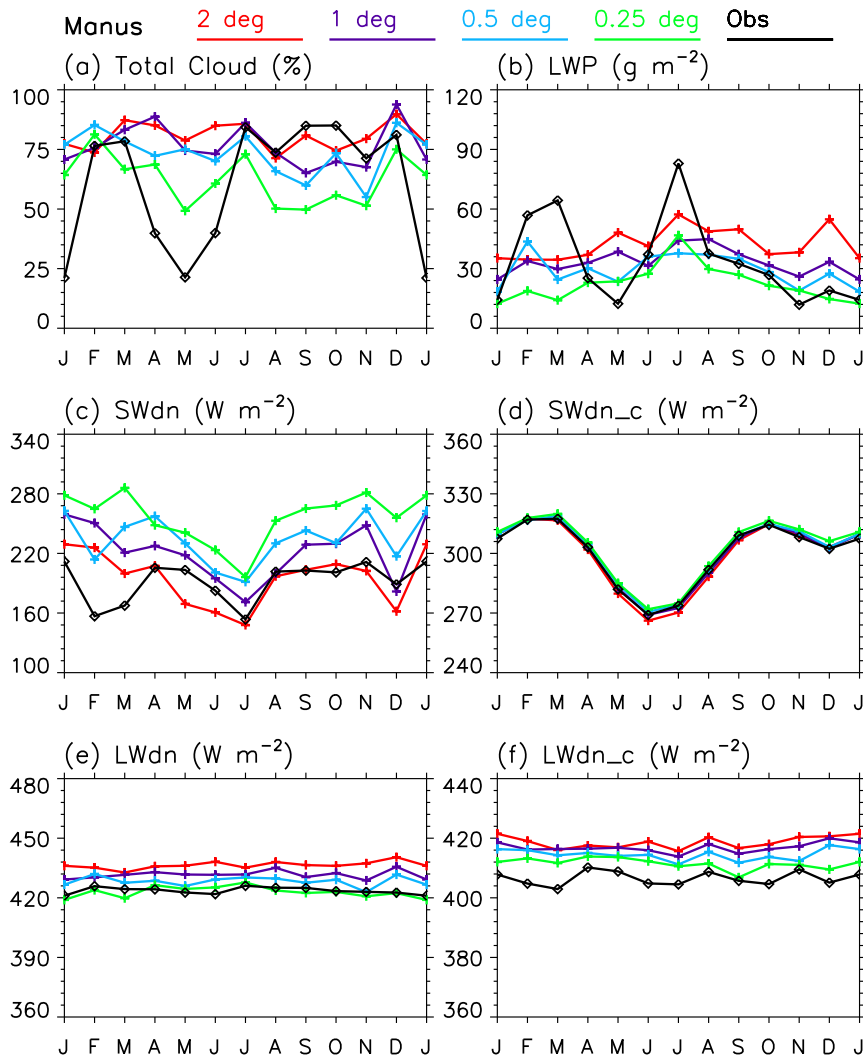
FIG. 1. Observed (black) and CAM5 simulated seasonal variations of (a) total cloud fraction, (b) LWP, (c) all-sky downwelling SW flux at the surface (SWdn), (d) clear-sky downwelling SW flux at the surface (SWdn_c), (e) LWdn, and (f) LWdn_c in 2009 at the ARM Manus site. The CAM5 was run at different horizontal grid spacings (2°, 1°, 0.5°, and 0.25°) using SD from the ECMWF analysis.

does give a much better estimate of cloud amount and SW radiative fluxes during the wet season (DJF), but the impact of model resolution is counterintuitive. We will look further into this in the following sections.

b. Vertical distribution of clouds

In section 3a we showed large model–observation discrepancies and sensitivity to model resolution in the two-dimensional total cloud fraction and other cloud-related quantities (Figs. 1–4 and Figs. S1 and S2). This section examines the vertical distribution of cloud fraction. Figure 5 shows the model–observation comparison and resolution dependency of the monthly mean vertical

profiles of total cloud fraction in 2009 at the Manus site. Total monthly cloud fraction in the observations is defined as the fraction of time with a given month that a cloud (of any type) was detected at a given height by the ARM vertically pointing active sensors. Total cloud fraction in a model grid denotes the horizontal fraction of the grid covered by all types of clouds within the grid box, including convective, stratiform, liquid, and ice clouds. Higher-resolution simulations generally give lower total cloud fraction at all levels, which is also true at the other two sites (not shown). At the Manus site where ground-based measurements are available, this resolution dependency in cloud fraction leads to a

TABLE 3. Annual mean cloud and radiation properties from ARM observations and model-grid-spacing simulations in 2009.

		TCF (%)	LWP (g m^{-2})	SWdn (W m^{-2})	SWdn_c (W m^{-2})	LWdn (W m^{-2})	LWdn_c (W m^{-2})
Manus	2°	80.7	43.1	192.8	298.7	436.5	418.5
	1°	76.9	34.0	219.2	299.8	431.8	416.7
	0.5°	73.3	30.0	232.4	300.8	428.4	414.4
	0.25°	62.2	23.1	255.1	302.1	423.2	411.5
	Obs	63.1	35.1	190.6	299.6	423.8	406.5
Nauru	2°	63.1	28.0	243.0	300.8	428.5	416.0
	1°	63.7	26.5	245.9	301.6	426.0	412.6
	0.5°	64.2	28.9	246.2	302.3	424.6	411.0
	0.25°	56.0	22.5	262.2	303.9	420.1	408.0
	Obs	—	—	232.7	302.3	424.9	412.1
Darwin	2°	49.3	26.1	240.5	292.6	410.3	402.2
	1°	55.4	77.9	218.4	292.1	409.7	396.6
	0.5°	54.0	38.9	231.8	289.5	415.4	404.6
	0.25°	50.2	39.5	239.0	290.1	413.4	403.3
	Obs	—	—	233.5	293.4	409.1	396.6

substantially lower bias for high clouds (above 10 km) but reduces the frequency of low clouds (below 5 km) too much in the 0.25° simulation (Fig. 5d).

As mentioned in the model description, the model calculates separate fractions for convective, liquid stratiform, and ice stratiform cloud types. We decompose the observed clouds into similar categories in order to diagnose potential deficiencies in the corresponding model parameterizations. Figure 6 shows long-term monthly mean frequencies of these cloud types using the observations and the 10-yr CAM5_2NG simulation. The model–observation discrepancy in total cloud fraction is similar to that in the 2009 comparison shown in Fig. 5, which can be attributed to the representation of specific cloud types. In both modeled and observed fields, the magnitude of cloud frequency above 5 km is dominated by stratiform clouds and by convective clouds below 5 km. The modeled seasonal cycle of convective cloud fraction, peaking in May–August, is out of phase with observations that have two separate peaks in late summer and winter, respectively. However, the overall magnitude and vertical extent of simulated convective clouds agree well with the observations at Manus. The seasonal cycle of cloud fraction is better simulated for stratiform clouds, but the vertical location is off by several kilometers and the magnitude of upper-tropospheric peak is largely biased. The model simulates a peak of liquid stratiform cloud fraction near the freezing level (~6 km), which does exist in the observations of total cloud fraction but not in the decomposed stratiform liquid clouds. This may be related to our simple separation of convective and stratiform clouds based only on their heights and depths. The thin layer of observed liquid stratiform clouds is likely formed through the melting of stratiform ice precipitation followed by cooling and increased stability (e.g., Johnson et al. 1996; Riihimaki et al. 2012). Satellite observations using lidar have revealed that such thin midlevel clouds are ubiquitous in the tropics and the magnitude of their radiative cooling effect could be as large as the warming effect of cirrus (e.g., Bourgeois et al. 2016). Biases in simulating this type of cloud may have a broader impact on the tropical branch of the large-scale circulations or heat and moisture transport as a result of cloud feedbacks.

The most apparent model biases in the total cloud fraction appear to be attributable to stratiform ice cloud, more specifically to its high frequency of overcast conditions (see appendix A). However, this could be partly due to the aforementioned observational limitation of ground-based active sensors. Following the filtering method used by Hogan et al. (2001) and Illingworth et al. (2007) to filter out potentially undetectable high clouds, we do see a discernable reduction in high-level cloud fraction, but the model still has a large overestimation of ice stratus (Fig. B2; see appendix B). Running the model at higher resolutions (e.g., the year 2009 simulations) does help to reduce such positive biases in the frequency of ice clouds, but also inadvertently increases the negative biases in the frequency of convective clouds and low-level liquid clouds (not shown). One plausible explanation for this is that convection in the higher-resolution simulations is less frequent, which would directly impact the convective cloud fraction and indirectly affect the free-tropospheric stratiform cloud frequency by either reducing the transport of water vapor out of the boundary layer or the detrainment of liquid and ice. This hypothesized drying effect severely reduces low-level liquid cloud frequency, but may improve the ice cloud simulation to some extent by offsetting the high relative humidity (RH) bias in the upper troposphere.
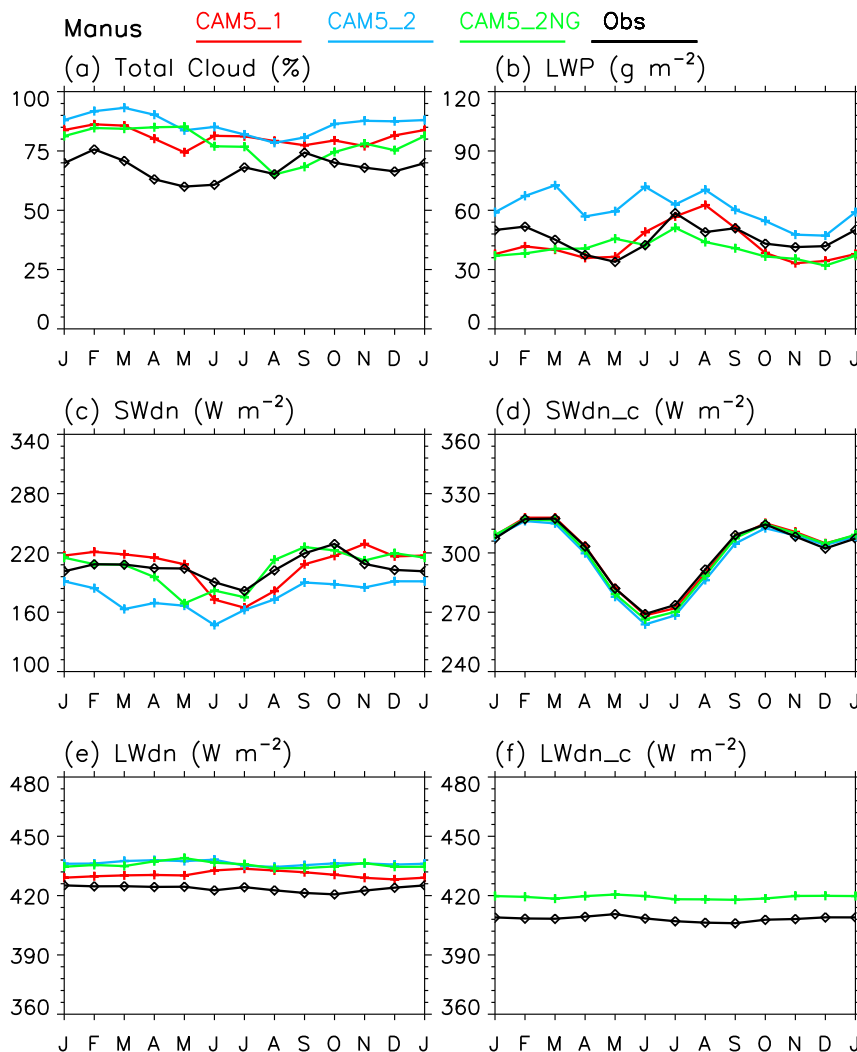
FIG. 2. As in Fig. 1, but for set 2 of the CAM5 simulations and the long-term observational datasets at the Manus site. CAM5_1 and CAM5_2 are FR simulations and CAM5_2NG is the simulation where the winds are nudged to the ECMWF reanalysis.

One facet of the simulations evident in both Figs. 5 and 6 is the excessive cloud layer near 1 km. This thin layer of persistent clouds is present in the various simulations regardless of resolution, running mode, or time period, indicating a deficiency in model representation of physical processes in the tropical boundary layer. While these lower-level (liquid) clouds have a large SW cloud radiative effect as a result of their high frequency of occurrence (e.g., Burleyson et al. 2015), the excessive ice clouds seem to have an insignificant impact on SW radiative fluxes. This may explain why the 2° simulation gives better overall performance in the estimate of surface SW fluxes compared to the finer-resolution simulations (Fig. 1 and Figs. S1 and S2) even if it has a much larger high-cloud fraction. Results regarding the

long-term simulations described above are roughly consistent among the three ARM TWP sites, with the exception that CAM5 can only capture the seasonal cycle of all cloud types at the Darwin site (not shown).

c. Role of meteorology

Results in the previous sections have shown that the nudged simulation can reproduce the multiyear mean seasonal cycles better than the free-running simulation at the same resolution (Figs. 2–4), but there are still significant biases in the frequency of clouds and the SW radiative fluxes at the surface. It is likely that the vertical distribution of moisture and/or temperature is still biased even with the nudging, which could also potentially explain some of the biases in downwelling LW fluxes

TABLE 4. Annual mean cloud and radiation properties from 10-yr ARM observations and model simulations.

		TCF (%)	LWP (g m^{-2})	SWdn (W m^{-2})	SWdn_c (W m^{-2})	LWdn (W m^{-2})	LWdn_c (W m^{-2})
Manus	CAM5_1	80.6	43.1	205.9	299.9	430.7	—
	CAM5_2	86.2	60.8	176.0	297.1	436.4	—
	CAM5_2NG	78.0	40.3	203.9	298.7	435.6	419.2
	Obs	67.7	45.5	205.2	299.6	423.5	408.2
Nauru	CAM5_1	67.3	27.9	243.7	301.4	424.0	—
	CAM5_2	72.1	39.0	230.6	299.2	430.7	—
	CAM5_2NG	62.5	26.7	247.8	301.4	426.5	414.0
	Obs	52.7	24.6	237.9	302.3	420.6	408.4
Darwin	CAM5_1	58.6	74.4	214.4	293.2	409.6	—
	CAM5_2	64.1	40.4	216.8	292.1	412.5	—
	CAM5_2NG	52.2	27.6	236.1	292.8	411.0	402.1
	Obs	66.1	32.9	232.4	293.5	407.1	394.7

and the excessive cloudiness at some levels. Figure 7 shows vertical profiles of the difference in annual mean moisture and temperature between the long-term model simulations (set 2 in Table 2) and the ARM soundings. The 10-yr free-running simulations (CAM5_1 and CAM5_2) show a more humid and cooler atmosphere throughout the lower troposphere compared to the ARM soundings. This bias occurs in all seasons (Fig. 8). The moisture bias in the free troposphere is reduced in the nudged simulation, in which a dry bias actually occurs in late summer and fall (Fig. 8f). The cold bias in the lower troposphere persists despite nudging, but the warm bias in the middle and upper troposphere changes to a cold bias in the nudged simulation.

The vertical profiles of the annual mean bias in moisture and temperature show an interesting layered structure with three distinct layers: the trade-wind boundary layer, a layer between the trade-wind inversion (~2 km) and the melting level (~6 km), and the layer up to the tropopause. The magnitude of bias varies among these layers. There is a shift in the temperature bias near the melting level in both the free-running and nudged or specified dynamics simulations (Fig. 9), indicating a change in the cause of the biases above and below this level that is likely related to model physics. The strong cold bias above the tropopause, which is likely a result of longwave radiative cooling induced by the excessive clouds underneath that layer (Fig. 6), is reduced but still present in the nudged simulation (Figs. 7b,e,h). In addition to the temperature biases, both the free-running and nudged simulations have an excess of moisture within the boundary layer and just above the freezing level (Figs. 7a,d,g). Too much moisture at this level could be the result of excessive detrainment of moisture lofted by penetrating convective clouds, a process included but poorly constrained in most convective parameterizations (e.g., Park and Bretherton 2009). As a result of the biases in moisture

and temperature, the model has a persistent positive bias in RH throughout the troposphere that is consistent between the free-running and nudged simulations (Figs. 7c,f,i). This could explain the persistent overestimation of stratiform clouds, especially in the liquid layer above freezing level and the ice layer in the upper troposphere. These vertical patterns are generally similar for all three sites (not shown).

The question is whether the remaining biases in moisture and humidity in the nudged simulation are due to the inadequate observational constraints compared to the model run using specified dynamics. To answer this, we further examine the meteorological fields in the specified dynamics simulations (Fig. 9). Even with the prescribed temperature, surface pressure, and heat and moisture fluxes from analysis, there are still significant differences in vertical profiles of temperature and RH between the specified dynamics simulations and the ARM soundings (Fig. 9). The structure of vertical layers shown in the profiles of moisture and temperature differences is similar to that in the nudged simulation (cf. Figs. 7 and 9). The temperature and moisture biases also appear to have a resolution dependency. The temperature bias is slightly worse in the higher-resolution simulations, where the mean bias is around 1.5 K in the lower troposphere and up to 3 K in the middle and upper troposphere in the 0.25° simulation. The excessive moisture in the low levels of the atmosphere, which was also identified as the cause of excessive low clouds near the mixed-layer top by Chandra et al. (2015) in their 1°-grid CAM5 simulations, remains in the higher-resolution simulations but is confined to the layer below 1 km (Figs. 9a,d,g). Moreover, there is a progressive transition from a moist bias to a dry bias in the free troposphere as the horizontal resolution increases, leading to large underestimates of RH all year round at Manus and Nauru and in the dry season at Darwin (not shown). This explains why clouds forming at 1–10 km
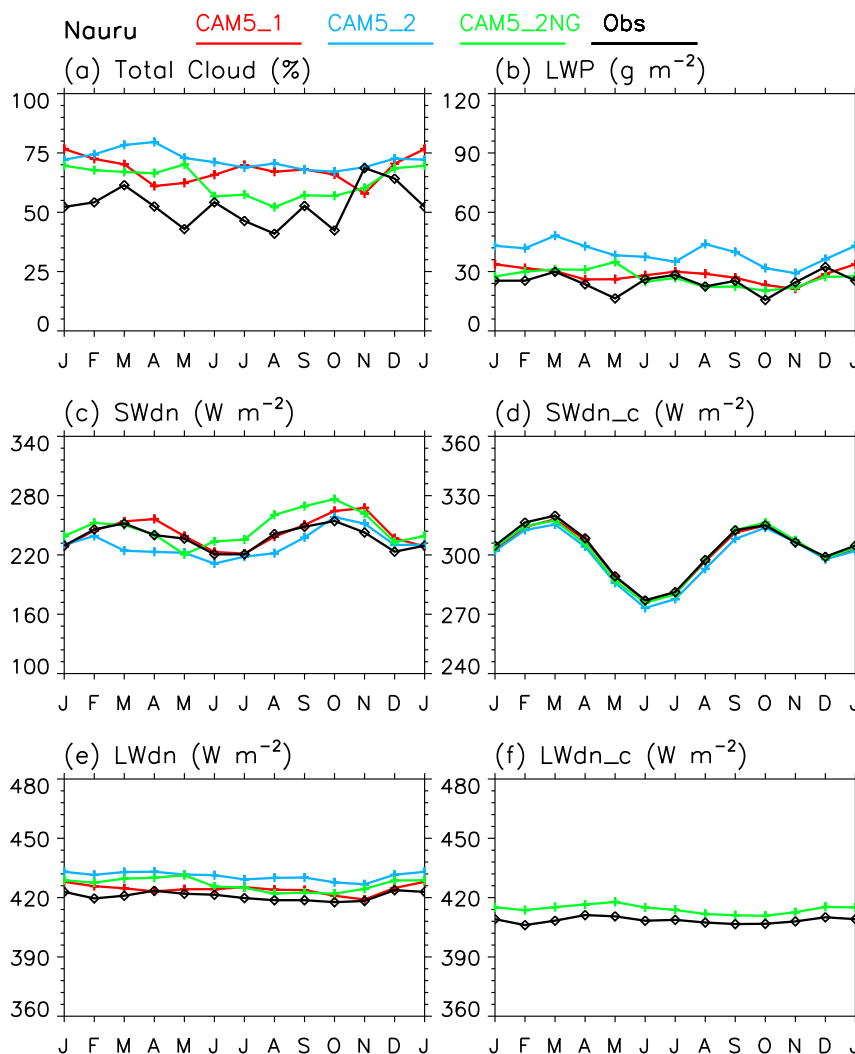
FIG. 3. As in Fig. 2, but for the Nauru site.

are underestimated in the 0.25° simulation (Fig. 5d). The fundamental cause of this drier midtroposphere in the higher-resolution simulations is probably rooted in the convection parameterization. We will explore this further in the next section.

d. Diurnal variations of clouds and radiation

High-frequency cloud and radiation fields from the 10-yr CAM5_2NG simulation as well as the long-term observations are sorted into eight 3-h bins to characterize the diurnal variations at each of the three TWP sites. Figure 10 shows the statistics of cloud properties and radiative fluxes in each time window across the diurnal cycle. The modeled and observed diurnal cycles of total cloud fraction are out of phase at all three sites, as is the diurnal cycle of LWP at Manus. The diurnal cycle of LWP is better simulated at Nauru and Darwin. At

Manus, CAM5 simulated overcast conditions for over 50% of the days that have measurements available (i.e., the median value of total cloud frequency is 100%). This gives an overall high bias in mean cloud fraction, but the mean LWP is lower compared to the observations, especially in the late afternoon. This is the likely cause of the positive bias in the afternoon surface downwelling SW fluxes (Fig. 10c). Mean total cloud fraction is also overestimated at Nauru, but underestimated at Darwin. The magnitude of LWP is well simulated at these two sites. Both all-sky and clear-sky LW fluxes are significantly overestimated in the model at Manus and Nauru (Figs. 10d,h). This is consistent with the positive biases seen in the seasonal variations (Figs. 2 and 3), which may be due to the inclusion of ocean into the model grid boxes containing the ARM sites. In addition, the model does not capture the observed diurnal
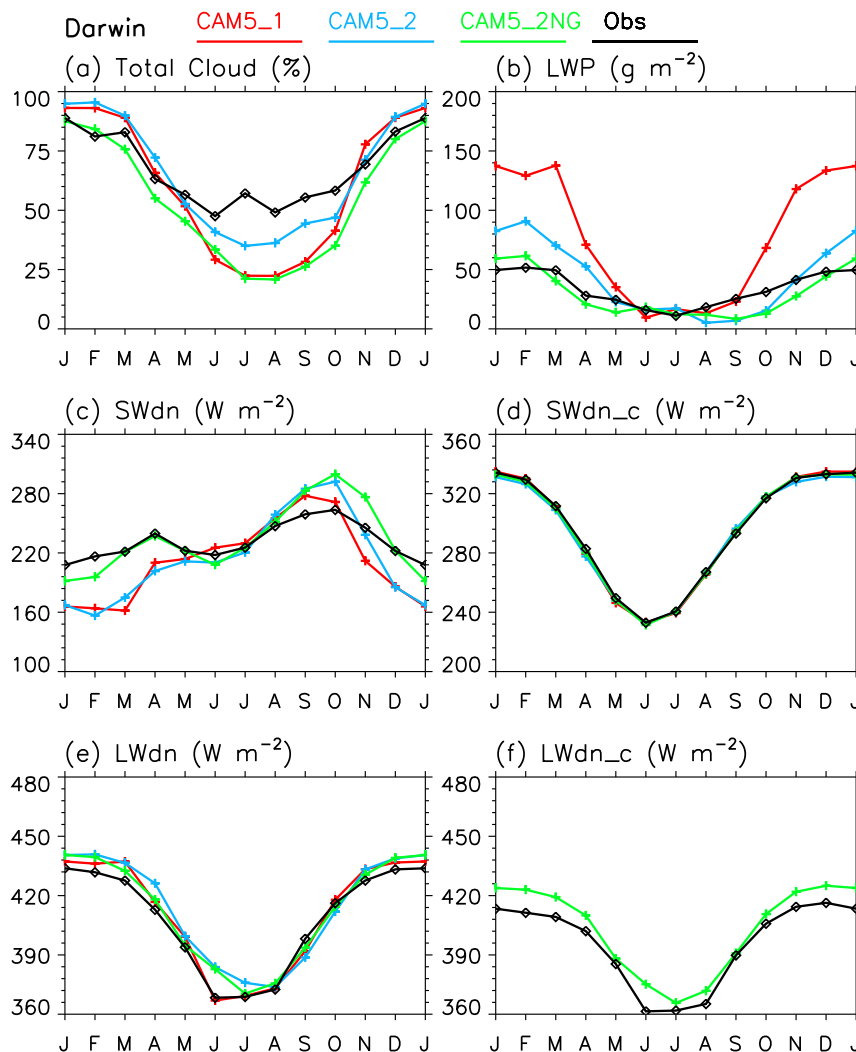
FIG. 4. As in Fig. 2, but for the Darwin site.

variation of LW fluxes (marked by a persistent suppression of LW fluxes overnight), indicating that the rapid adjustment of surface temperature by radiative cooling over the island after the sun sets cannot be captured by the mean state of the model grid box that covered by a mixture of land and water. Longwave fluxes are better simulated at the Darwin site, where the spread (i.e., the day-to-day variability) is much larger compared to the other two sites (Fig. 10i). This is likely due to the strong seasonality and moisture contrast in the ambient conditions associated with monsoon circulations in and around the Darwin ARM site. Further separation of the cloud and radiation properties by season at Darwin reveals that the spread mainly occurs in the dry season (JJA), when both cloud fraction and LWP are also significantly smaller (not shown). In the wet season (DJF), the model captures the frequent

overcast conditions at Darwin. The diurnal variations (and the spread in each 3-h bin) of the cloud and radiation quantities are better captured by the model in both seasons compared to our analysis of the annual means, suggesting that model evaluation against measurements at Darwin is better done separately for the wet and dry seasons, which echoes the finding of Nguyen et al. (2015). Neither the observed nor modeled diurnal variations differ significantly by season at Manus and Nauru.

To investigate the bias in the diurnal variations of total cloud fraction we examine the diurnal cycles of cloud frequency by cloud type (Fig. 11; see also Figs. S3 and S4 in the supplemental material). Observational cloud fractions in this analysis are defined similarly to the monthly values (i.e., the fraction of time that a given cloud type was present at a given height within each 3-h window). As with the two-dimensional total cloud
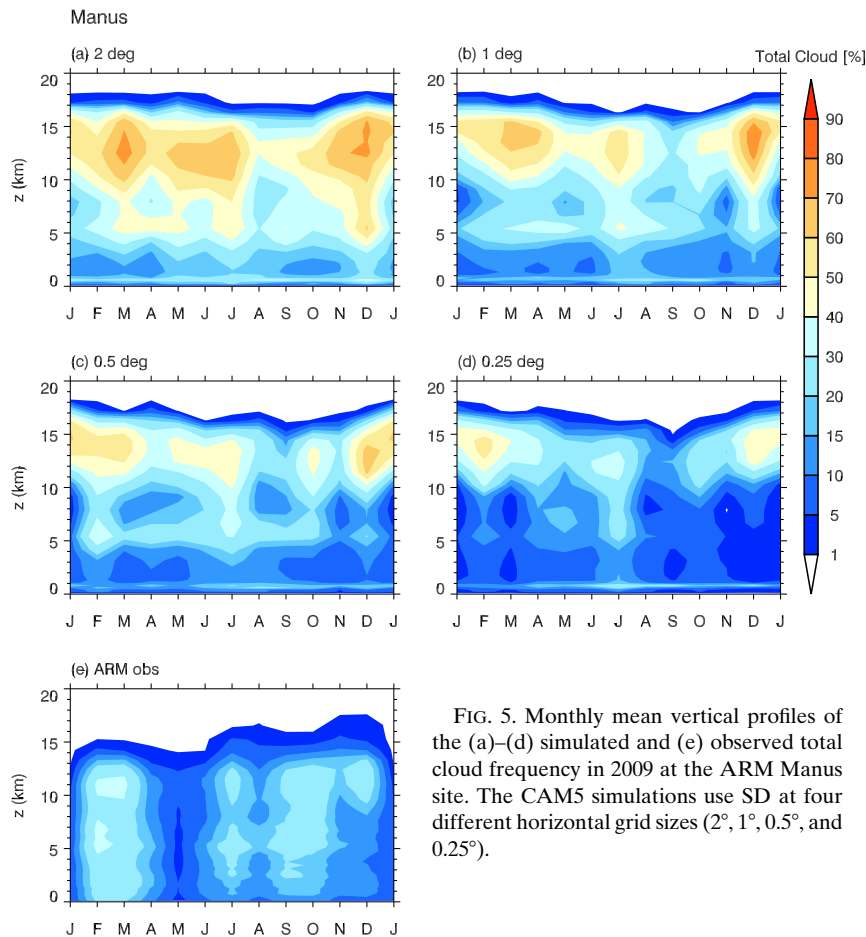
Manus

FIG. 5. Monthly mean vertical profiles of the (a)–(d) simulated and (e) observed total cloud frequency in 2009 at the ARM Manus site. The CAM5 simulations use SD at four different horizontal grid sizes (2°, 1°, 0.5°, and 0.25°).

fractions, both CAM5 and the observations show a distinct diurnal cycle in the vertical distribution of total cloud fraction at all three sites. However, the modeled and observed diurnal time–height plots are out of phase and have large differences in the magnitude, especially at Manus. The total cloud frequency bias appears to be dominated by the modeled stratiform ice clouds, which have a maximum frequency before dawn (approximately 1800–2100 UTC; Figs. 11e and S3e). This also represents the peak CAM5 total cloud fraction (Figs. 11a and S3a). Modeled convective clouds also have a distinct diurnal cycle, which has similar amplitude but is out of phase with the observations. Although the convective clouds occur relatively infrequently, they likely dominate the diurnal variation of LWP and daily mean SW cloud radiative effect. The diurnal cycle of cloud fraction for both modeled (which peaks from 1500 to 2100 UTC; Figs. 11c and S3c) and observed (which peaks from 0300 to 0600 UTC; Fig. 11d) convective clouds matches their corresponding diurnal cycle of LWP (Fig. 10b). This is consistent with the observational analysis in Burleyson et al. (2015) and indicates that potential deficiencies in

the phase of convection predicted by a model's cumulus parameterization can have an important impact on the local energy budget by biasing the SW cloud radiative effects at the surface.

The midlevel liquid stratiform cloud fraction is overestimated in the model and lacks a distinct diurnal cycle, a feature that is similar to the observations. These midlevel liquid clouds have a very small amount of water, making them unobservable by the ground-based remote sensing instruments (Figs. B1 and B2; see appendix B), which might partly explain the model overestimation. However, the overall high bias in the liquid and ice stratiform cloud fraction is more likely related to the overestimate of RH, as shown in the model–observation comparison of mean thermodynamic profiles (Figs. 7 and 8). The timing of the modeled peak in supercooled liquid and ice cloud frequency underneath the tropopause aligns well with the peak time of deep convection (cf. Fig. 11i with Fig. 11c and Fig. S3i with Fig. S3c), suggesting that transport of water vapor or the detrainment of liquid and ice from deep convection is a plausible cause of the excessive stratiform ice cloud frequency in the upper troposphere.
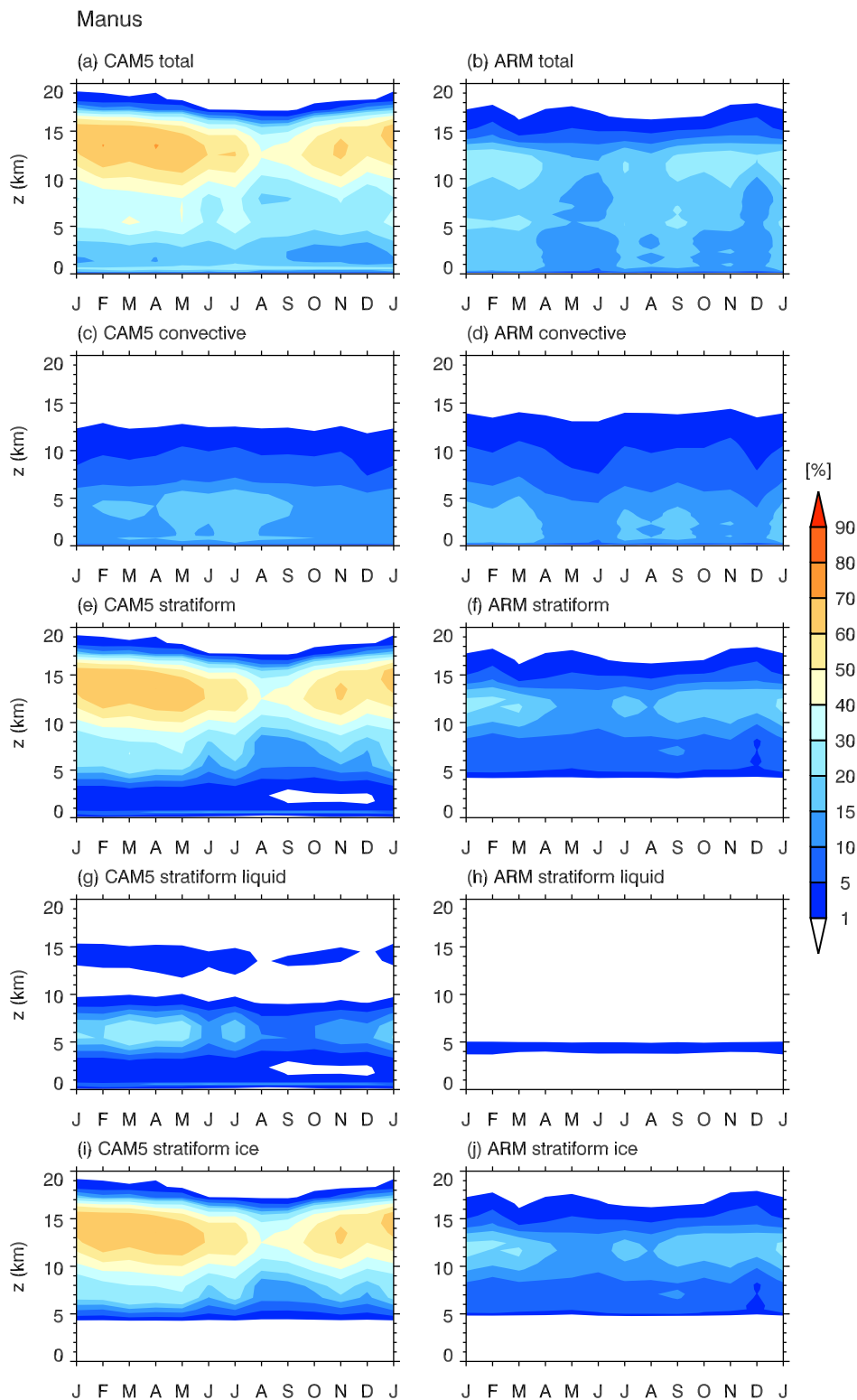
Manus

FIG. 6. Monthly mean vertical profiles of cloud frequency by cloud type in (a),(e),(g),(i) the 10-yr CAM5_2NG simulation and (b),(f),(h),(j) the observed long-term averages at the ARM Manus site.
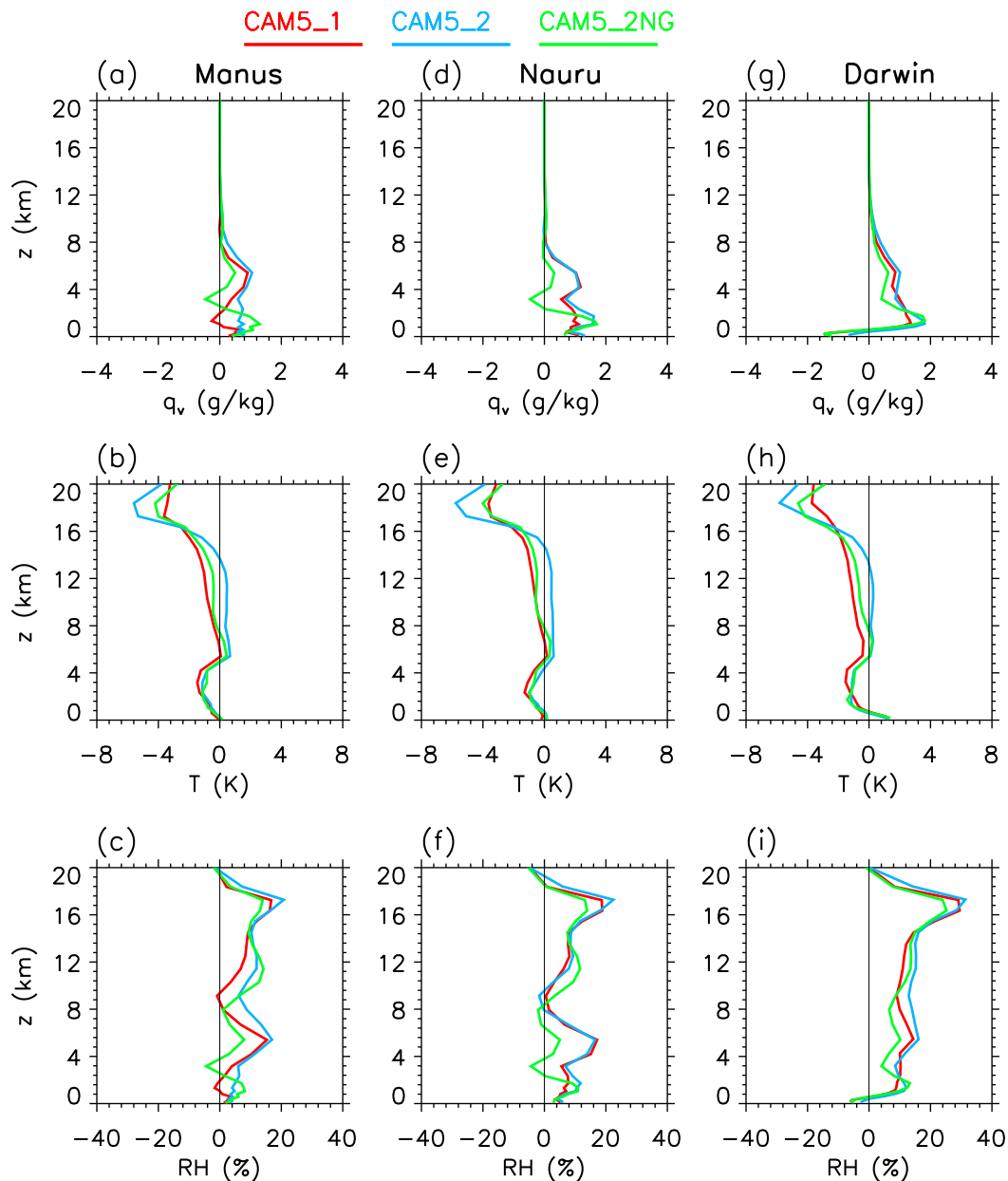
FIG. 7. Vertical profiles of the mean difference in (a),(d),(g) water vapor mixing ratio q_v, (b),(e),(h) temperature T, and (c),(f),(i) RH between the three simulations in set 2 (CAM5_1, CAM5_2, and CAM5_2NG in red, blue, and green, respectively) and the long-term observations at the (left) Manus, (center) Nauru, and (right) Darwin ARM sites.

4. Summary and discussion

Climate models have large biases in predicting clouds and their radiative effects over the tropics. The long-term ARM observational datasets collected at the three TWP sites provide an excellent resource for evaluating climate models using both statistical and process-oriented approaches. The diversity of simultaneous measurements made by ARM also allows for a more complete investigation of biases in multiple model components. In this study, we use the ARM datasets to evaluate the ability of CAM5 to simulate the various types of clouds, their seasonal and diurnal variations, and their impact on surface radiation at the ARM TWP sites. Our approach can be easily adapted for the evaluation of other global or regional model simulations and can be used to track the performance of climate models over time as they evolve to include new treatment of dynamics and physics.
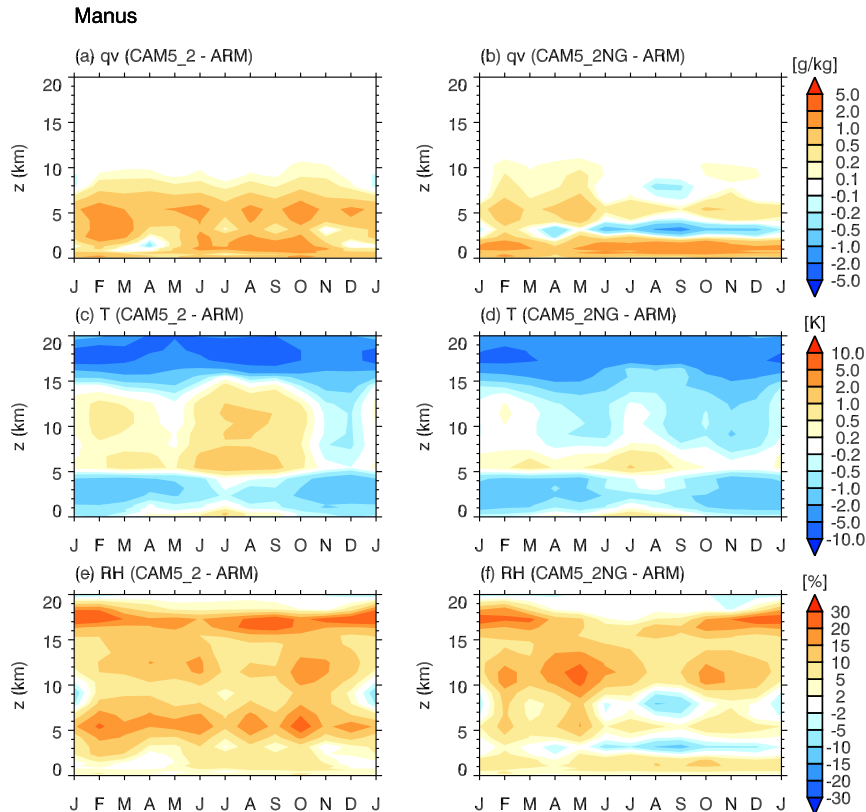
Manus

FIG. 8. Monthly mean differences in (a),(b) q_v, (c),(d) T, and (e),(f) RH between the (left) CAM5_2 and (right) CAM5_2NG simulations and the long-term observations at the ARM Manus site.

We conducted two sets of CAM5 simulations at various horizontal grid sizes (around 2°, 1°, 0.5°, and 0.25°) to study the resolution dependency of modeled clouds and their radiative effects. The modeled meteorological fields were constrained by analysis or reanalysis values (in nudging or SD mode). Note that the definition of cloud types in CAM5 cloud parameterizations is different from that used to derive cloud types from the observational products, so rather than attempting to draw exact quantitative conclusions from the model–observation comparison of specific cloud types we focus on identifying likely sources of the qualitative biases that are readily apparent.

The most apparent model bias in total cloud fraction is attributable to the large overestimation of stratiform ice clouds. However, this may be partly due to an underestimation of the frequency of high clouds caused by observational constraints of the ground-based instruments. As described in Burleyson et al. (2015), the underestimation of high clouds in these tropical ARM datasets is due to a combination of decreasing sensitivity with increasing height, interference from the background solar radiation, and the frequent presence of

dense liquid clouds in the column. Hogan et al. (2001) derived a vertical profile of minimum detectable ice water content (IWC) to filter out unobservable thin cirrus cloud in models. Using this filtering method, Illingworth et al. (2007) documented changes in midlatitude high cloud fraction on the order of −10%. A recent study using satellite data to bound the error estimates found that the ARM lidar at Darwin underestimates the frequency of high clouds by roughly 25% (Thorsen et al. 2013). The magnitude of reduction in stratiform ice cloud fraction for one of our simulations estimated using the filtering method is similar to those in the previous studies (Fig. B2). Thus we believe the model biases in stratiform ice cloud fraction shown in this present study are larger than the observational uncertainty. Bias correction using the filtering method is not expected to affect our conclusion. Even with a cloud radar simulator applied in a climate model that is close to the CAM5 used in our study, Zhang et al. (2018) showed a large underestimate of the occurrence of low clouds and a large overestimate (more than twice as frequent) of high clouds over the southern Great Plains. This is consistent with our findings. Another cause of the
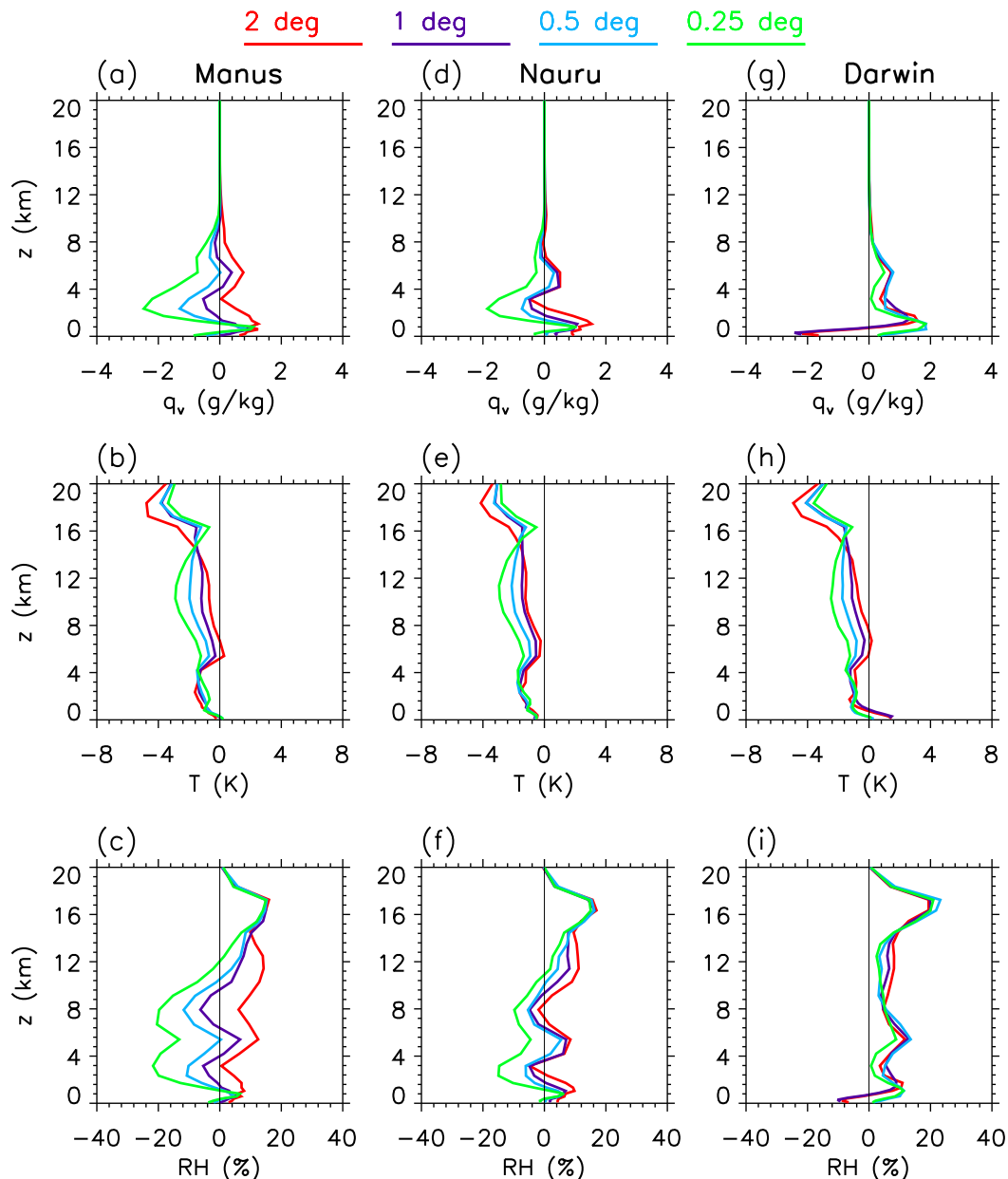
FIG. 9. As in Fig. 7, but for the differences between the CAM5 SD simulations (at four different horizontal resolutions) and the ARM observations in 2009 at the three ARM sites.

bias in total clouds is the negative bias in temperature and the positive bias in humidity in the model, which affects clouds at all levels. Our results show that even with strong meteorological constraints from analysis/reanalysis, the model simulations still have significant differences in their vertical profiles of temperature and humidity compared to the ARM soundings. This may also explain why the model does not capture the interannual variability in the seasonal cycles of clouds that is masked by the biases in meteorological fields.

Model biases in the seasonal cycle of total cloudiness are weakly dependent on model resolution. The 0.25° simulation has the lowest bias in annual mean total cloud fraction, but the model reproduces the all-sky downwelling SW flux better at coarser resolutions. The model misses the seasonal cycles at all resolutions. A relatively stronger resolution dependence is found in the frequency for specific types of clouds, indicating a lack of scale awareness in the CAM5 cloud parameterizations. Higher-resolution simulations do reduce positive biases
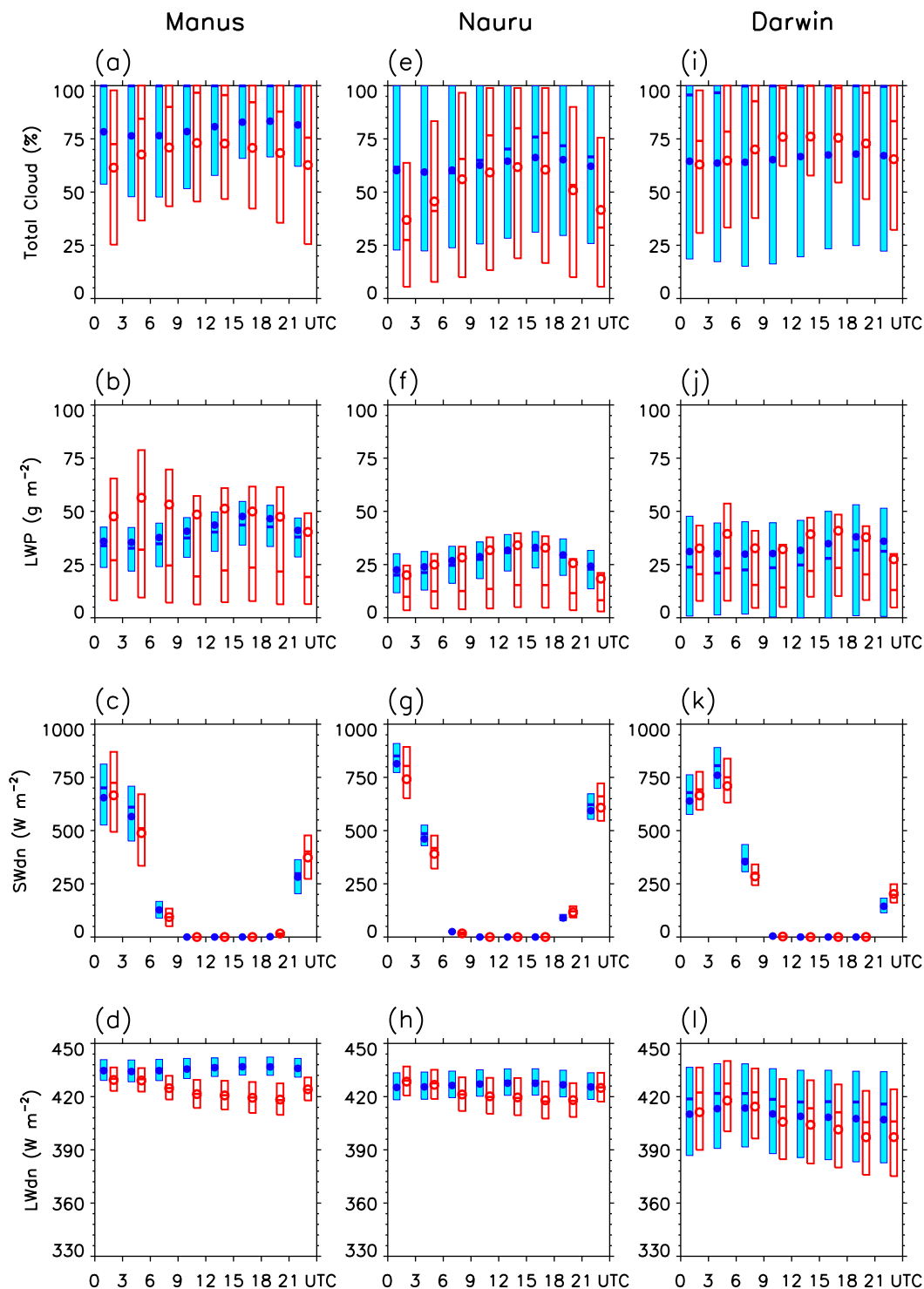
FIG. 10. Distributions across the diurnal cycle of the CAM5_2NG (blue) and ARM observed (red) cloud properties and radiative fluxes at the (a)–(d) Manus, (e)–(h) Nauru, and (i)–(l) Darwin ARM sites. Boxes denote the 25th, 50th, and 75th percentiles and the mean values are marked with a circle. Statistics for each 3-h window are based on all available samples (one per day) at each site. Days with missing observational data are removed from model time series for consistency.
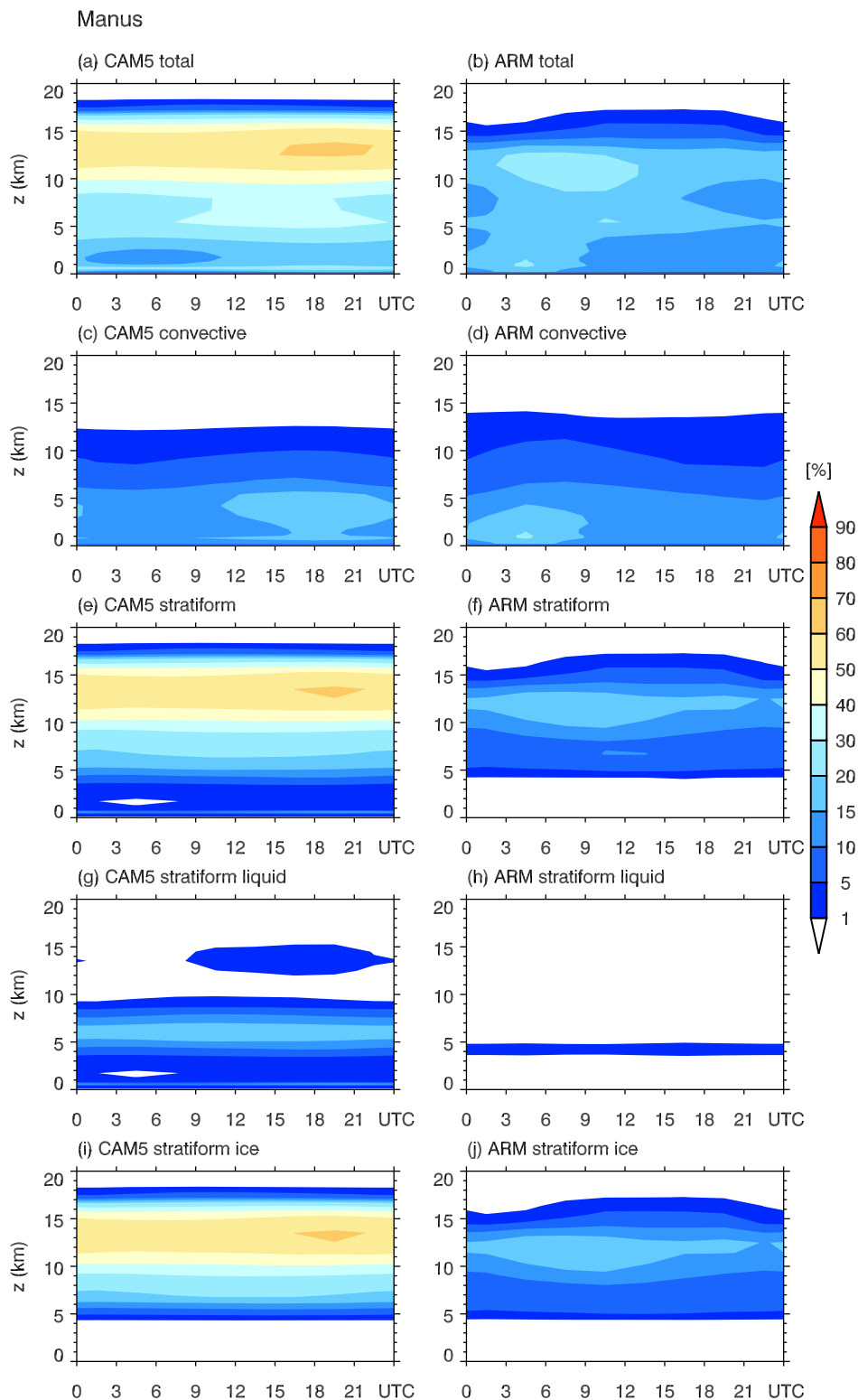
Manus

FIG. 11. The 3-h mean vertical profiles of cloud frequency by cloud type in the (a),(c),(e),(g),(i) 10-yr CAM5_2NG simulation and (b),(d),(f),(h),(j) observed long-term averages at the ARM Manus site.

in the frequency of ice clouds, but they inadvertently increase the negative biases in convective clouds and lower-level liquid clouds. One plausible explanation for this is that convection in the higher-resolution simulations is weaker or less frequent, which directly impacts the convective cloud fraction and indirectly affects the stratiform clouds by redistributing water vapor in the vertical. Biases in clouds can further feed back onto meteorology. Model biases in the profiles of temperature and humidity (and RH) also have a resolution dependency that is consistent with the potential problem in resolution-dependent performance of convection parameterizations mentioned above. Note that CAM5 diagnoses the cloud fraction of liquid and ice stratus using grid-mean RH, critical RH, and an assumed distribution function, where constant critical RH values are set for ice stratus and low-level and high-level liquid stratus, respectively (Park et al. 2014). Quaas (2012) found distinct geographical distributions of critical RH from a combination of satellite data and reanalysis products, suggesting that these RH parameters should be linked to model resolved dynamics. This can potentially make the cloud parameterizations scale-aware.

The excessive ice clouds in CAM5 seem to have an insignificant impact on the surface SW radiative fluxes, which appear to be more sensitive to biases in the frequency of mid- and low-level liquid clouds. A significant positive bias is found in the modeled all-sky and clear-sky LW downwelling fluxes. This can be partly explained by the mismatch between point measurements on islands and the model gridbox mean that is determined by the model-simulated atmospheric and cloud properties over the warmer sea surface. The bias is smaller during the day than at night when islands cool off. Finer-resolution simulations do have smaller biases in LW fluxes.

Both CAM5 and the observations show a distinct diurnal cycle in total cloud fraction, but the model and the observations have large differences in the amplitude of the diurnal cycle and tend to be out of phase. Modeled convective clouds have a distinct diurnal cycle, which has a similar magnitude but is out of phase with the observed diurnal cycle of convective clouds. Although convective cloud fractions are relatively small, convective clouds likely dominate the variations in LWP and SW cloud radiative effects. Thus any deficiencies in model cumulus parameterizations that affect the frequency or characteristics of convective clouds can have a significant impact on the total cloud radiative effect in the model. Our analysis on the frequency of occurrence of convective cloud fraction (see appendix A) confirms the model deficiency in triggering deep convection. There was a well-known deficiency in the original deep convection scheme (Zhang and McFarlane 1995) that uses a threshold value of

convective available potential energy (CAPE) to trigger deep convection (e.g., Xie and Zhang 2000). Many studies over the years have attempted to improve the representation of deep convection. A revised Zhang–McFarlane scheme using CAPE calculated with a dilute entraining plume to trigger convection (e.g., Neale et al. 2008), which has been included in the CAM5 version here, is still insufficient to bring the model in agreement with the observations. However, several recent developmental representations of deep convection (e.g., Wang et al. 2016; Chen and Mapes 2018) do dramatically affect the frequency of deep convection in CAM5.

This deficiency in deep convection could also manifest itself in the out-of-phase diurnal cycle. On the other hand, our analysis suggests that the out-of-phase deep convection affects the vertical distribution and diurnal cycle of stratiform clouds through the transport of vapor or the detrainment of liquid and ice, pointing to the importance of cumulus parameterizations in determining the model performance of simulating both convective and stratiform clouds.

Model biases in the diurnal cycle of some cloud properties and radiative fluxes vary by site. For example, at Manus and Nauru a positive bias in cloud cover occurs year round during the daytime, but for the Darwin site more noteworthy negative biases tend to occur at night. The diurnal cycles at Darwin are better captured by the model in both the dry and wet seasons compared to the annual means, suggesting that model evaluation against measurements at Darwin is better done separately by season. Neither the observed nor modeled diurnal variations show a discernable seasonal dependence at Manus or Nauru.

Some results from our analysis may be used to guide future model–observation comparisons. Caution is warranted when using point measurements over an island to evaluate grid mean quantities in global models because subgrid-scale land–ocean proportions within the model grid box are unresolved. More specifically, the CAM5 has excessive surface LW fluxes compared to site measurements when the grid is partially covered with ocean that is expected to be warmer than land at night. We also found that it can be hard to separate cause and effect by just looking at model biases because of the coupling between clouds, temperature, and moisture. For example, excessive liquid clouds at a given level may cool the layer below and heat the layer above via interactions with the ambient radiation. These heating and cooling effects could in turn lead to biases in cloud frequency at adjacent levels of the model. The melting of excessive ice precipitation followed by cooling and increased stability may lead to the formation of thin liquid layers that have important radiative cooling
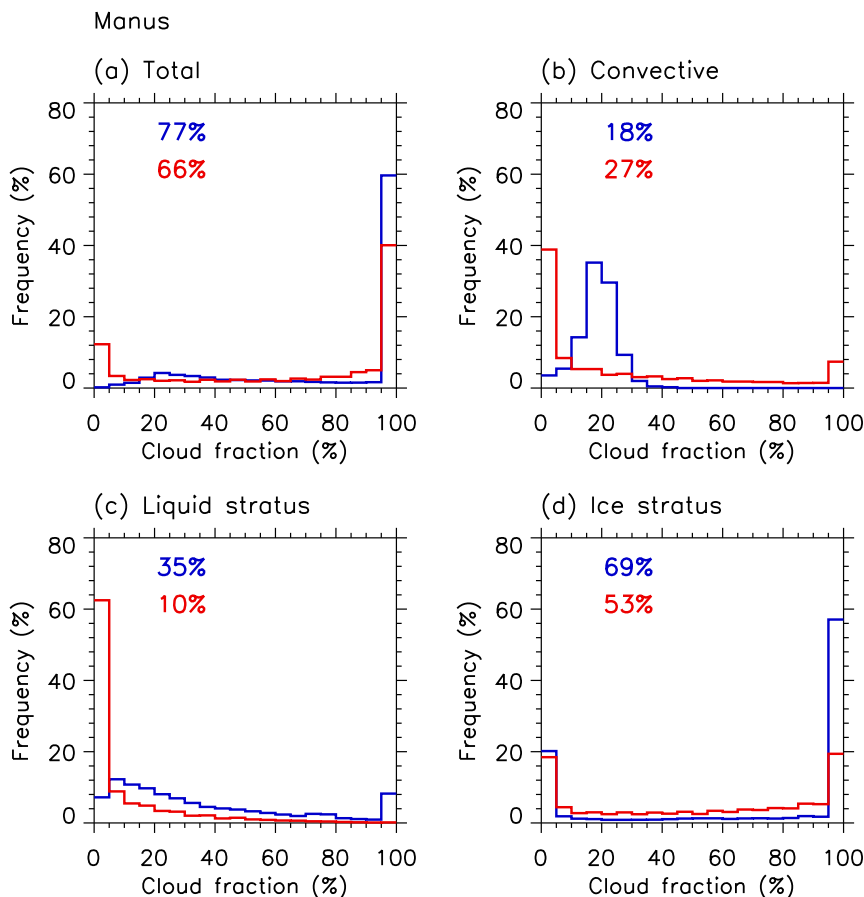
FIG. A1. Frequency of occurrence (%) of 3-hourly (a) total, (b) convective, (c) liquid stratus, and (d) ice stratus cloud fraction over the Manus site from the 10-yr CAM5_2NG experiment (blue) and ARM observations (red). The numbers in the legend are the corresponding mean cloud fraction for the experiment and observations.

effect and even a broader impact on the tropical branch of the large-scale circulations. One final outcome of our study was that analyses of the root causes of biases in simulated diurnal cycles are hampered by the limited temporal sampling of sounding observations, which are most often only collected twice a day. Thus the observations are largely unable to characterize the diurnal variations in meteorological profiles that are necessary to help identify the causes of model biases.

With increasing computational power, climate models are achieving higher and higher resolutions, but have not yet reached global cumulus-scale resolutions (~100 m). Thus the representation of cumulus convection in next-generation climate models will still rely on parameterization schemes that use a variety of assumptions to relate subgrid-scale quantities to grid-scale variables (e.g., Randall et al. 2003; Arakawa 2004). Cumulus parameterization schemes are responsible for most of the difficulties many climate models have in representing the right ratios of shallow and deep clouds,

the transition from shallow to deep convection, and the appropriate sensitivity of convection to tropospheric humidity. These difficulties are partly due to the separate parameterization of shallow and deep convection. Unified and/or scale-aware parameterizations are being developed for current and next-generation climate models to address these and other shortcomings. Our approach can be used to evaluate the behavior of those new parameterizations at higher spatial resolution to see if they are any better than the coarse results shown here.
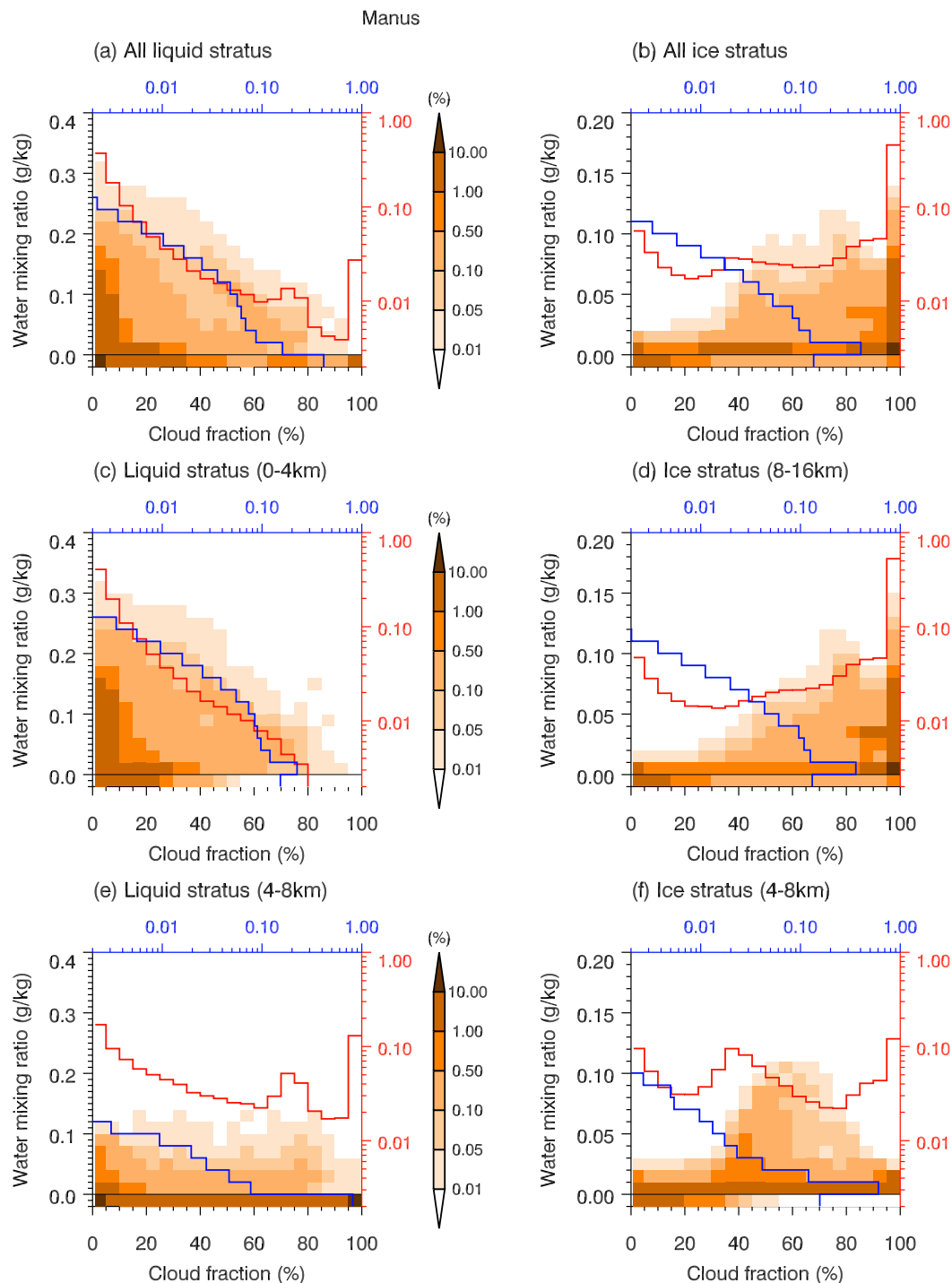
FIG. B1. Joint histogram (shading) of cloud fraction (greater than 1%) and water mixing ratio for (a) all liquid stratus, (b) all ice stratus, (c) liquid stratus (below 4 km), (d) ice stratus (8–16 km), (e) liquid stratus (4–8 km), and (f) ice stratus (4–8 km) over the Manus site based on 3-hourly mean cloud fields (experiment CAM5_2NG). The red line is the individual histogram for cloud fraction (right y axis) and the blue line for cloud water mixing ratio (top x axis). Cloud samples (individual model layers within the altitude range) with in-cloud water mixing ratio smaller than the threshold value (2×10^{-4} g kg^{-1}) are identified as empty stratus and placed in the bottommost bin.
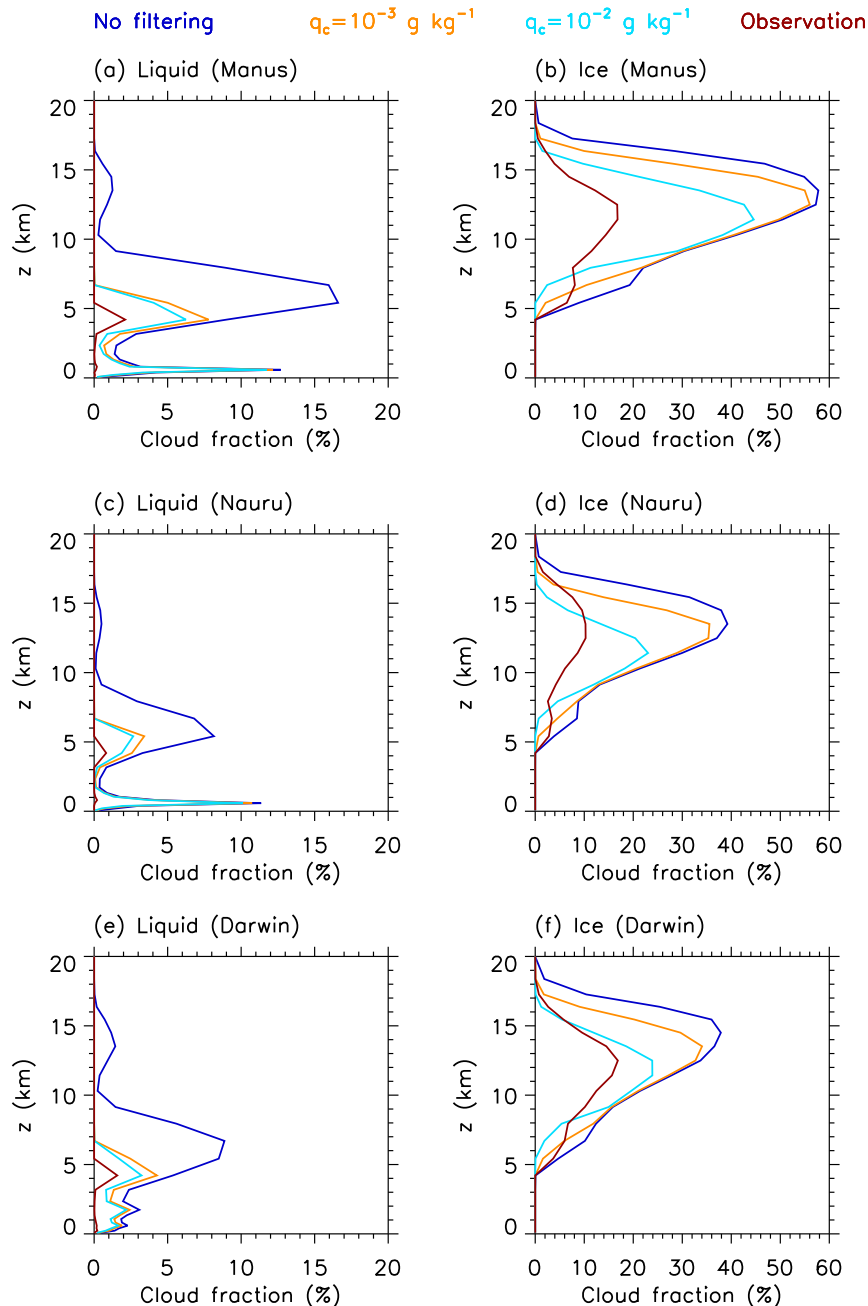
FIG. B2. Vertical profiles of liquid stratus and ice stratus over (a),(b) Manus, (c),(d) Nauru, and (e),(f) Darwin based on 3-hourly mean cloud fields (blue; experiment CAM5_2NG). Light blue and orange lines represent modeled cloud fraction after the filtering process with threshold values of 10^{-2} and 10^{-3} g kg^{-1}, respectively.

APPENDIX A

Frequency of Occurrence of 3-Hourly Mean Cloud Fraction

When comparing modeled mean cloud fraction with observations, we did not distinguish between "cloud

frequency of occurrence'' and ''mean cloud amount when cloud is present.'' Sometimes the cancellation of biases between the two can make the modeled mean cloud fraction look better than it actually is (e.g., Morcrette et al. 2012). For convective clouds, the two are governed by different parts of the parameterization schemes. To help attribute model errors to individual parameters or formulas, we analyzed the frequency of occurrence of 3-hourly mean cloud fields from the 10-yr CAM5_2NG experiment and ARM observations at the Manus site (Fig. A1). The model overestimates the frequency of overcast (>95%) liquid and ice stratus. The modeled occurrence of clear-sky conditions (<5%) is less frequent than the observed ones for total, liquid stratus, and convective clouds. Comparison of the modeled and observed frequency distributions suggests that convective clouds are produced too often in the model and the amount is too small when cloud is present. For the CAM5 deep convection scheme (Zhang and McFarlane 1995), the overestimated frequency has been attributed to the use of CAPE in the triggering function (e.g., Xie and Zhang 2000; Neale et al. 2008), while the cloud amount is largely determined by several tunable parameters in the scheme (e.g., Yang et al. 2013).

APPENDIX B

The Empty Stratus Issue in CAM5

The use of diagnostic cloud fraction and prognostic microphysical schemes for stratiform clouds in CAM5 can cause inconsistencies between the stratus fraction and the in-cloud condensate in stratus clouds. Although there are inconsistency checks to remove ''empty'' stratus at the end of stratus macrophysics step, stratus can still become empty after the subsequent microphysics due to precipitation (Park et al. 2014). To help assess the model biases in cloud fraction, we did an analysis of the joint probability distributions of cloud fraction and water mixing ratio for both liquid and ice stratus clouds at various altitude ranges (Fig. B1). The analysis is based on 3-hourly average cloud fields in the 10-yr CAM5_2NG simulation using a minimum in-cloud water mixing ratio of $2 \times 10^{-4}\,\mathrm{g\,kg^{-1}}$ as the threshold to identify empty status (Park et al. 2014). The frequency of empty liquid stratus integrated over all cloud fraction bins is over 40%, primarily contributed by midlevel liquid stratus (4–8 km). Liquid stratus forming below 4 km in the model have only 22% empty clouds. Empty ice stratus is much less frequent, around 10% integrated over all height levels.

These empty or thin clouds are likely to be undetectable by ground-based remote sensors. Hogan et al. (2001) proposed to use a minimum detectable ice water threshold to filter out unobservable thin cirrus clouds in models. To get an estimate on the impact of such filtering (i.e., setting thin clouds to zero) on our results, two threshold values of cloud water mixing ratio, 10^{-3} and $10^{-2}\,\mathrm{g\,kg^{-1}}$, are used to recalculate the vertical distribution of liquid and ice stratus cloud fraction (Fig. B2), where the latter one is considered to provide the lower bound of detectable cloud fractions. The midlevel liquid clouds do show a strong reduction in response to the filtering, which might explain some of the model overestimation. Cloud fractions at certain heights are still largely overestimated in the model even with the strong filtering ($10^{-2}\,\mathrm{g\,kg^{-1}}$), especially for the ice stratus clouds at all three ARM sites.

REFERENCES

Ackerman, T. P., W. E. Clements, F. J. Barnes, and D. S. Renné, 1999: Tropical Western Pacific cloud and radiation testbed: Science, siting, and implementation strategies, ARM Tech. Rep. ARM-99-004, 71 pp., www.arm.gov/publications/site_reports/twp/arm-99-004.pdf.

Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. J. Climate, 17, 2493–2525, https://doi.org/10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2.

Bony, S., and J. Dufresne, 2005: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. Geophys. Res. Lett., 32, L20806, https://doi.org/10.1029/2005GL023851.

Bourgeois, Q., A. M. Ekman, M. R. Igel, and R. Krejci, 2016: Ubiquity and impact of thin mid-level clouds in the tropics. Nat. Commun., 7, 12432, https://doi.org/10.1038/ncomms12432.

Boyle, J. S., 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, https://doi.org/10.1029/2004JD005042.

Bretherton, C. S., and S. Park, 2009: A new moist turbulence parameterization in the Community Atmosphere Model. J. Climate, 22, 3422–3448, https://doi.org/10.1175/2008JCLI2556.1.

Burleyson, C., C. Long, and J. Comstock, 2015: Quantifying diurnal cloud radiative effects by cloud type in the tropical western Pacific. J. Appl. Meteor. Climatol., 54, 1297–1312, https://doi.org/10.1175/JAMC-D-14-0288.1.

Chandra, A. S., C. Zhang, S. A. Klein, and H.-Y. Ma, 2015: Low-cloud characteristics over the tropical western Pacific from ARM observations and CAM5 simulations. J. Geophys. Res. Atmos., 120, 8953–8970, https://doi.org/10.1002/2015JD023369.

Chen, B., and B. E. Mapes, 2018: Effects of a simple convective organization scheme in a two-plume GCM. J. Adv. Model. Earth Syst., https://doi.org/10.1002/2017MS001106, in press.

Clothiaux, E. E., and Coauthors, 2001: The ARM Millimeter Wave Cloud Radars (MMCRs) and the Active Remote Sensing of Clouds (ARSCL) Value Added Product (VAP). DOE Tech. Memo. ARM VAP-002.1, 56 pp., https://www.arm.gov/publications/tech_reports/arm-vap-002-1.pdf.

Comstock, J. M., and C. Jakob, 2004: Evaluation of tropical cirrus cloud properties derived from ECMWF model output and

ground based measurements over Nauru Island. *Geophys. Res. Lett.*, **31**, L10106, https://doi.org/10.1029/2004GL019539.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Hinkelman, L. M., T. P. Ackerman, and R. T. Marchand, 1999: An evaluation of NCEP Eta model predictions of surface energy budget and cloud properties by comparison with measured ARM data. *J. Geophys. Res.*, **104**, 19 535–19 549, https://doi.org/10.1029/1999JD900120.

Hogan, R. J., C. Jakob, and A. J. Illingworth, 2001: Comparison of ECMWF winter-season cloud fraction with radar-derived values. *J. Appl. Meteor.*, **40**, 513–525, https://doi.org/10.1175/1520-0450(2001)040<0513:COEWSC>2.0.CO;2.

Hurrell, J. W., and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Illingworth, A. J., and Coauthors, 2007: CloudNet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Amer. Meteor. Soc.*, **88**, 883–898, https://doi.org/10.1175/BAMS-88-6-883.

Jakob, C., G. Tselioudis, and T. Hume, 2005: The radiative, cloud, and thermodynamic properties of the major tropical western Pacific cloud regimes. *J. Climate*, **18**, 1203–1215, https://doi.org/10.1175/JCLI3326.1.

Johnson, R., P. Ciesielski, and K. Hart, 1996: Tropical inversions near the 0°C level. *J. Atmos. Sci.*, **53**, 1838–1855, https://doi.org/10.1175/1520-0469(1996)053<1838:TINTL>2.0.CO;2.

Lin, G., H. Wan, K. Zhang, Y. Qian, and S. J. Ghan, 2016: Can nudging be used to quantify model sensitivities in precipitation and cloud forcing? *J. Adv. Model. Earth Syst.*, **8**, 1073–1091, https://doi.org/10.1002/2016MS000659.

Long, C. N., and T. P. Ackerman, 2000: Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *J. Geophys. Res.*, **105**, 15 609–15 626, https://doi.org/10.1029/2000JD900077.

——, and D. D. Turner, 2008: A method for continuous estimation of clear-sky downwelling longwave radiative flux developed using ARM surface measurements. *J. Geophys. Res.*, **113**, D18206, https://doi.org/10.1029/2008JD009936.

——, and Coauthors, 2013: ARM research in the equatorial western Pacific: A decade and counting. *Bull. Amer. Meteor. Soc.*, **94**, 695–708, https://doi.org/10.1175/BAMS-D-11-00137.1.

Ma, P.-L., and Coauthors, 2013: The role of circulation features on black carbon transport into the Arctic in the Community Atmosphere Model version 5 (CAM5). *J. Geophys. Res. Atmos.*, **118**, 4657–4669, https://doi.org/10.1002/jgrd.50411.

——, and Coauthors, 2015: How does increasing horizontal resolution in a global climate model improve the simulation of aerosol–cloud interactions? *Geophys. Res. Lett.*, **42**, 5058–5065, https://doi.org/10.1002/2015GL064183.

Madden, R. A., and P. R. Julian, 1994: Observations of the 40–50-day tropical oscillation—A review. *Mon. Wea. Rev.*, **122**, 814–837, https://doi.org/10.1175/1520-0493(1994)122<0814:OOTDTO>2.0.CO;2.

Mather, J. H., T. P. Ackerman, W. E. Clements, F. J. Barnes, M. D. Ivey, L. D. Hatfield, and R. M. Reynolds, 1998: An atmospheric radiation and cloud station in the tropical western Pacific. *Bull. Amer. Meteor. Soc.*, **79**, 627–642, https://doi.org/10.1175/1520-0477(1998)079<0627:AARACS>2.0.CO;2.

McFarlane, S. A., C. N. Long, and J. Flaherty, 2013: A climatology of surface radiative effects at the ARM tropical western Pacific sites. *J. Appl. Meteor. Climatol.*, **52**, 996–1013, https://doi.org/10.1175/JAMC-D-12-0189.1.

Medeiros, B., B. Stevens, I. Held, M. Zhao, D. Williamson, J. Olson, and C. Bretherton, 2008: Aquaplanets, climate sensitivity, and low clouds. *J. Climate*, **21**, 4974–4991, https://doi.org/10.1175/2008JCLI1995.1.

Morcrette, C. J., E. J. O'Connor, and J. C. Petch, 2012: Evaluation of two cloud parametrization schemes using ARM and Cloud-Net observations. *Quart. J. Roy. Meteor. Soc.*, **138**, 964–979, https://doi.org/10.1002/qj.969.

Morrison, H., and A. Gettelman, 2008: A new two-moment bulk stratiform microphysics scheme in the Community Atmosphere Model, version 3 (CAM3). Part I: Description and numerical tests. *J. Climate*, **21**, 3642–3659, https://doi.org/10.1175/2008JCLI2105.1.

Nam, C., S. Bony, J.-L. Dufresne, and H. Chepfer, 2012: The 'too few, too bright' tropical low-cloud problem in CMIP5 models. *Geophys. Res. Lett.*, **39**, L21801, https://doi.org/10.1029/2012GL053421.

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, https://doi.org/10.1175/2008JCLI2244.1.

——, and Coauthors, 2012: Description of the NCAR Community Atmosphere Model (CAM 5.0). NCAR Tech. Note NCAR/TN-486+STR, 274 pp., http://www.cesm.ucar.edu/models/cesm1.0/cam/docs/description/cam5_desc.pdf.

Nguyen, H., A. Protat, V. Kumar, S. Rauniyar, M. Whimpey, and L. Rikus, 2015: A regional forecast model evaluation of statistical rainfall properties using the CPOL radar observations in different precipitation regimes over Darwin, Australia. *Quart. J. Roy. Meteor. Soc.*, **141**, 2337–2349, https://doi.org/10.1002/qj.2525.

Paquin-Ricard, D., C. Jones, and P. A. Vaillancourt, 2010: Using ARM observations to evaluate cloud and clear-sky radiation processes as simulated by the Canadian regional climate model GEM. *Mon. Wea. Rev.*, **138**, 818–838, https://doi.org/10.1175/2009MWR2745.1.

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, https://doi.org/10.1175/2008JCLI2557.1.

——, ——, and P. J. Rasch, 2014: Integrating cloud processes in the Community Atmosphere Model, version 5. *J. Climate*, **27**, 6821–6856, https://doi.org/10.1175/JCLI-D-14-00087.1.

Qian, Y., C. N. Long, H. Wang, J. M. Comstock, S. A. McFarlane, and S. Xie, 2012: Evaluation of cloud fraction and its radiative effect simulated by IPCC AR4 global models against ARM surface observations. *Atmos. Chem. Phys.*, **12**, 1785–1810, https://doi.org/10.5194/acp-12-1785-2012.

Quaas, J., 2012: Evaluating the "critical relative humidity" as a measure of subgrid-scale variability of humidity in general circulation model cloud cover parameterizations using satellite data. *J. Geophys. Res.*, **117**, D09208, https://doi.org/10.1029/2012JD017495.

Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.*, **84**, 1547–1564, https://doi.org/10.1175/BAMS-84-11-1547.

Richter, J. H., and P. J. Rasch, 2008: Effects of convective momentum transport on the atmospheric circulation in the Community Atmosphere Model, version 3. *J. Climate*, **21**, 1487–1499, https://doi.org/10.1175/2007JCLI1789.1.

Riihimaki, L. D., S. A. McFarlane, and J. M. Comstock, 2012: Climatology and formation of tropical midlevel clouds at the Darwin ARM site. *J. Climate*, **25**, 6835–6850, https://doi.org/10.1175/JCLI-D-11-00599.1.

Thorsen, T. J., Q. Fu, J. M. Comstock, C. Sivaraman, M. A. Vaughan, D. M. Winker, and D. D. Turner, 2013: Macrophysical properties of tropical cirrus clouds from the CALIPSO satellite and from ground-based micropulse and Raman lidars. *J. Geophys. Res.*, **118**, 9209–9220, https://doi.org/10.1002/jgrd.50691.

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, https://doi.org/10.1175/2011BAMS3095.1.

Wang, H., and W. Su, 2013: Evaluating and understanding top of the atmosphere cloud radiative effects in Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) Coupled Model Intercomparison Project phase 5 (CMIP5) models using satellite observations. *J. Geophys. Res. Atmos.*, **118**, 683–699, https://doi.org/10.1029/2012JD018619.

Wang, Y., G. J. Zhang, and G. C. Craig, 2016: Stochastic convective parameterization improving the simulation of tropical precipitation variability in the NCAR CAM5. *Geophys. Res. Lett.*, **43**, 6612–6619, https://doi.org/10.1002/2016GL069818.

Xie, S., and M. Zhang, 2000: Impact of the convection triggering function on single-column model simulations. *J. Geophys. Res.*, **105**, 14 983–14 996, https://doi.org/10.1029/2000JD900170.

——, and Coauthors, 2010: Clouds and more: ARM climate modeling best estimate data. *Bull. Amer. Meteor. Soc.*, **91**, 13–20, https://doi.org/10.1175/2009BAMS2891.1.

Yang, B., and Coauthors, 2013: Uncertainty quantification and parameter tuning in the CAM5 Zhang–McFarlane convection scheme and impact of improved convection on the global circulation and climate. *J. Geophys. Res. Atmos.*, **118**, 395–415, https://doi.org/10.1029/2012JD018213.

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, https://doi.org/10.1080/07055900.1995.9649539.

Zhang, K., and Coauthors, 2014: Technical note: On the use of nudging for aerosol–climate model intercomparison studies. *Atmos. Chem. Phys.*, **14**, 8631–8645, https://doi.org/10.5194/acp-14-8631-2014.

Zhang, M. H., and Coauthors, 2005: Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *J. Geophys. Res.*, **110**, D15S02, https://doi.org/10.1029/2004JD005021.

Zhang, Y., and Coauthors, 2018: The ARM cloud radar simulator for global climate models: Bridging field data and climate models. *Bull. Amer. Meteor. Soc.*, **99**, 21–26, https://doi.org/10.1175/BAMS-D-16-0258.1.