ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 34, MARCH 2017, 306–320

• **Original Paper** •

Evaluation of NASA GISS Post-CMIP5 Single Column Model Simulated Clouds and Precipitation Using ARM Southern Great Plains Observations

Lei ZHANG[1,2], Xiquan DONG[*1,3], Aaron KENNEDY[3], Baike XI[3], and Zhanqing LI[2]

[1]*College of Global Change and Earth System Science, Beijing Normal University, Beijing* 100875*, China*
[2]*Department of Atmospheric and Oceanic Sciences, University of Maryland, College Park, Maryland* 20740*, USA*
[3]*Department of Atmospheric Sciences, University of North Dakota, Grand Forks, North Dakota* 58203*, USA*

(Received 2 December 2015; revised 5 July 2016; accepted 30 September 2016)

ABSTRACT

The planetary boundary layer turbulence and moist convection parameterizations have been modified recently in the NASA Goddard Institute for Space Studies (GISS) Model E2 atmospheric general circulation model (GCM; post-CMIP5, hereafter P5). In this study, single column model (SCM_P5) simulated cloud fractions (CFs), cloud liquid water paths (LWPs) and precipitation were compared with Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) ground-based observations made during the period 2002–08. CMIP5 SCM simulations and GCM outputs over the ARM SGP region were also used in the comparison to identify whether the causes of cloud and precipitation biases resulted from either the physical parameterization or the dynamic scheme. The comparison showed that the CMIP5 SCM has difficulties in simulating the vertical structure and seasonal variation of low-level clouds. The new scheme implemented in the turbulence parameterization led to significantly improved cloud simulations in P5. It was found that the SCM is sensitive to the relaxation time scale. When the relaxation time increased from 3 to 24 h, SCM_P5-simulated CFs and LWPs showed a moderate increase (10%–20%) but precipitation increased significantly (56%), which agreed better with observations despite the less accurate atmospheric state. Annual averages among the GCM and SCM simulations were almost the same, but their respective seasonal variations were out of phase. This suggests that the same physical cloud parameterization can generate similar statistical results over a long time period, but different dynamics drive the differences in seasonal variations. This study can potentially provide guidance for the further development of the GISS model.

Key words: single column model, model evaluation, cloud fraction, turbulence parameterization

Citation: Zhang, L., X. Q. Dong, A. Kennedy, B. K. Xi, and Z. Q. Li, 2017: Evaluation of NASA GISS post-CMIP5 single column model simulated clouds and precipitation using ARM Southern Great Plains observations. *Adv. Atmos. Sci.*, **34**(3), 306–320, doi: 10.1007/s00376-016-5254-4.

1. Introduction

Although many improvements have been made in the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor, 2001; Jiang et al., 2012; Klein et al., 2013; Lauer and Hamilton, 2013; Li et al., 2013; Wang and Su, 2013), cloud amount, distribution, diurnal cycle, and their relation to large-scale dynamics are still problematically simulated in climate models (e.g., Zhang et al., 2005; Qian et al., 2012; Su et al., 2013; Stanfield et al., 2014; Dolinar et al., 2015; Jiang et al., 2015) and numerical weather prediction models such as the NOAA global forecast system (Yoo and Li, 2012; Yoo et al., 2013). Motivated by recently changed planetary boundary layer (PBL) turbulence and moist convection parameterizations in the NASA Goddard Institute for Space Studies (GISS) Model E2 atmospheric general circulation model (GCM) (post-CMIP5, hereafter P5), Stanfield et al. (2014, 2015) investigated P5-simulated clouds and top-of-the-atmosphere (TOA) radiation budgets and found that P5 was significantly improved compared to its CMIP5 predecessor (C5). However, these studies were done on a global scale and without a detailed examination of the vertical distributions and diurnal variations of clouds. Improvements in the representation of regional clouds and related processes as a result of updated parameterizations were also not investigated in their studies.

Due to the complexities of evaluating GCMs with observations, the single column model (SCM) approach was developed to evaluate and test model parameterizations (Randall et al., 1996). It has been implemented by the Atmospheric Radiation Measurement (ARM) program (Ackerman and Stokes, 2003) to improve the representation of clouds and radiation in GCMs using long-term surface observations (Klein and Del

* Corresponding author: Xiquan DONG
Email: xdong@email.arizona.edu

Genio, 2006). SCMs take advantage of the fact that parameterizations for processes such as those that create clouds are independent for each column in a GCM. To run an SCM, it is necessary to generate forcing data to drive it. Forcing data used by others were derived from field experiment data or enhanced soundings obtained during intensive observing periods. These kinds of SCM studies (e.g., Ghan et al., 2000; Betts and Jakob, 2002; Xu et al., 2005; Xie et al., 2005) were limited to relatively short periods, usually from one week to one month.

Combining ARM long-term observations and the Rapid Update Cycle 2 weather forecast model (RUC), Xie et al. (2004) developed a continuous forcing product (1999–2001) over the ARM Southern Great Plains (SGP) site, which provides a tool for modelers to run SCMs over a longer time period. Kennedy et al. (2010) ran the GISS SCM (SI2000 version) using the three-year ARM forcing dataset and compared simulated cloud fractions (CFs) with ARM ground-based lidar data and Geostationary Operational Environmental Satellite observations. Using the same forcing data, Song et al. (2013, 2014) ran seven SCMs (including the GISS SCM) and compared simulated precipitation and CFs with ARM SGP observations. These studies demonstrated the importance of doing long-term SCM evaluations to diagnose the causes of cloud and precipitation biases in the models. Because an SCM is a single grid box of a GCM and lacks interactions with large-scale circulation and dynamic feedback present in the GCM, it would be ideal to compare both GCM and SCM outputs over the same grid box with observations. This comparison would help to determine whether the cloud and precipitation biases in the models result from model physical parameterizations or model dynamics because both the GCM and the SCM have the same physical parameterizations.

The SCM may accumulate large errors or depart too far away from observations when the running period is longer than a week. Therefore, either a reset or a relaxation technique is often used in SCM model simulations (Lohmann et al., 1999; Randall and Cripe, 1999; Hack and Pedretti, 2000; Iacobellis et al., 2000; Kennedy et al., 2010). Note that using the relaxation technique can result in unrealistically high or low precipitation in different model simulations. In addition, it is difficult to find a relaxation time scale that would keep the model's atmospheric states close to the observations and allow the model to develop its own process. In previous evaluation studies, different relaxation times were used, and model responses were varied (Kennedy et al., 2010; Song et al., 2014), which encouraged us to investigate the model sensitivity to the relaxation time. Therefore, multiple SCM runs were carried out using different relaxation time scales using the ARM SGP long-term forcing dataset. The simulated results were then compared with long-term ARM SGP merged soundings and cloud observations to achieve a better understanding of the response of SCM simulations to the relaxation time scale.

The primary purpose of this study is to evaluate the GISS Post-CMIP5 SCM (SCM_P5) simulated cloud and precipitation outputs using ARM SGP ground-based radar-lidar observations made during the period 2002–08. Seasonal and diurnal variations in CF and precipitation, as well as their vertical distributions, were examined. To further explore the causes of cloud and precipitation biases in the model simulations, the outputs over the same grid box from the GCM and the CMIP5 SCM (SCM_C5) were compared. Through this comparison, we can identify if the biases resulted from model physical parameterizations or model dynamics. The secondary goal of this paper is to test the model response to different relaxation time scales. This paper is structured as follows: Section 2 provides a brief description of the model, evaluation data, and methodology used in this study. Section 3 evaluates the SCM and GCM outputs against observations and investigates the possible causes for model biases. A summary of findings and conclusions are provided in section 4.

2. Datasets and methodology

2.1. *GISS-E2 SCM and GCM*

The GISS SCM_P5 used in this study is identical to the P5 GCM described by Stanfield et al. (2014, 2015) with updated PBL turbulence and moist convection parameterizations. The model has a horizontal resolution of $2^\circ \times 2.5^\circ$ latitude by longitude with 40 vertical layers, following sigma coordinates to 150 hPa with constant pressure layers above and a model top at 0.1 hPa. The dynamics and physics in the model are calculated and output every 30 minutes (Schmidt et al., 2014). P5 atmospheric GCM intermediate diagnostic data, which run with prescribed sea surface temperatures, are provided by the NASA GISS. C5 GCM simulations are obtained from the Earth System Grid Federation Program for Climate Model Diagnosis and Intercomparison database.

The GISS SCM uses a diagnostic cloud scheme that partitions CF into convective and stratiform clouds. The total CF is the sum of stratiform and convective CFs. In the stratiform parameterization, the model diagnoses large-scale CFs using a relative humidity (RH)–based scheme (Sundqvist, 1978; Sundqvist et al., 1989). A relationship between grid box mean RH and cloud occurrence is assumed. To form clouds in the model, a critical RH (U_{00}) must be reached. In the early version of the SCM (SI2000), U_{00} was set to 60% and held constant both in space and height. In the later version, the U_{00} is allowed to vary to some degree. For example, U_{00} is lowered in regions of upward motion and is also allowed to vary in the boundary layer by consideration of the variance in subgrid-scale moisture. In the latest version of P5, instead of specifying U_{00}, two parameters, U_a and U_b, are used to achieve radiative balance and a better cloud climatology (Yao and Cheng, 2012). U_a is primarily used to form clouds above 850 hPa, while U_b is for clouds below 850 hPa. In this study, U_a and U_b in the P5 (C5) model are set to 0.65 and 1.0, respectively, which are tuned to radiative balance. In addition, wmui_mulitplier is 2.0 (critical ice cloud water multiplier) and entrainment_cont1 is 0.4 (constant for entrainment rate).

The P5 version with the newer turbulence parameteri-

zation scheme differs from the C5 version in the following ways: (1) the diffusivities and the counter-gradient flux term in the nonlocal vertical transport scheme are modeled differently; (2) the turbulence length scale is defined differently; and (3) for the calculation of the PBL height, the P5 scheme uses the "Richardson number criterion", while C5 uses the "TKE criterion" as detailed by Yao and Cheng (2012). For the moist convection parameterization, the CF is determined by the updraft mass flux and the convective updraft speed (Del Genio and Yao, 1993). The P5 cumulus parameterization has been modified with increased entrainment and rain evaporation, and changes in the convective downdraft, as discussed by Del Genio et al. (2012).

2.2. *ARM continuous forcing*

To drive SCMs, atmospheric state variables, including temperature and humidity profiles, vertical and advective tendencies, must be specified. The ARM large-scale continuous forcing dataset over the ARM SGP site developed by Xie et al. (2004) is used in this study. This forcing dataset was produced through a combination of atmospheric state variables from the RUC 2 weather forecast model (Benjamin et al., 2004) and surface and TOA observations made at the ARM SGP site (Zhang et al., 1997). This forcing corresponds to a circular area approximately 180 km in radius centered on the SGP site (36.6°N, 97.5°W). The forcing data were originally developed for 1999–2001 from 40-km RUC-2 simulations, with an extension to the year 2008. Compared to the original three-year forcing dataset, the new forcing dataset (2002–08) has an updated physical parameterization with a higher resolution (13 km).

To be consistent, the new forcing dataset (June 2002–December 2008) was used to drive the SCM in this study. Forcing data from January–May 2002 and February 2003 are not available. Some parameters from the forcing, such as atmospheric temperature, RH profiles, precipitation rate, and liquid water path (LWP), are used as a ground truth to evaluate the modeled results at a temporal resolution of one hour.

2.3. *ARM ground-based observations*

The ground-based observations are primarily made by the ARM 35-GHz millimeter wavelength cloud radar (MMCR) at the ARM SGP site, which has a minimum detectable reflectivity factor of −55 dBZ at 1 km and −35 dBZ at 10 km (Moran et al., 1998). The MMCR operates at a wavelength of 8 mm in a vertically pointing mode and provides continuous profiles of radar reflectivity from hydrometeors moving through the radar field of view, allowing for the identification of clear and cloudy conditions. The beam width is 0.2° resulting in a horizontal resolution of ~40 m at 12 km above ground level. Belfort laser ceilometer and micropulse lidar (MPL) measurements are also used as an additional data source to determine the cloud base height (Clothiaux et al., 2000). Inclusion of laser ceilometer and MPL measurements allows for the filtering out of insects present near the ground during the spring and summer seasons at the SGP site. The total CF is the ratio of cloudy samples detected by the ARM radar-lidar to the total number of samples when both radar and lidar-ceilometer instruments were in operation (Xi et al., 2010). The high, middle, and low-level CFs are defined as >6 km, 3–6 km and <3 km, respectively. To address the comparability of CFs between models and ground observations, an extension of the ISCCP simulator for radar data is used (Del Genio et al., 2005). The simulator assumes a random-maximum cloud overlap assumption for stratiform clouds and maximally overlapped assumption for convective clouds, which are consistent with the treatment in the GISS model.

2.4. *Relaxation technique*

To prevent model errors from accumulating over time, the relaxation technique can be applied to relax the temperature and humidity profiles towards observations in a certain relaxation time scale. The use of relaxation does not hide model problems (Randall and Cripe, 1999) and can suppress the sensitivity to initial conditions (Hack and Pedretti, 2000). However, precipitation is always an issue in the SCM when using relaxation because spuriously high or low precipitation can occur, depending on the model. Song et al. (2013) found that the GISS SCM-produced precipitation amount was much smaller than observations when using a relaxation time scale of three hours. Kennedy et al. (2010) reported that the GISS SCM-simulated precipitation amount was close to the observed value at the SGP site when using the reset method every 24 h instead of relaxation.

2.5. *Simulation design*

To fully explore differences in the C5 and P5 model versions, and their response to the relaxation time scale, several model experiments were designed (Table 1). These simulations were all driven by the same ARM forcing data. The first set of P5 runs was relaxed on temperature, T, and specific humidity, q, at a time scale of three hours and is called the "SCM_P5 (3h)" run. In the second set of runs, a 24-h relaxation time scale was set for both T and q fields and is called the "SCM_P5 (24h)" run. For the third set of runs, the C5 SCM version was run with the relaxation time set at 24 h and is called the "SCM_C5 (24h)" run. Other P5 simulations with 6- and 12-h relaxation time are shown in Table 1. In this study, we mainly focus on the 3- and 24-h relaxation time; therefore, other relaxation times are not discussed in detail. All simulations were initialized at 0000 UTC at the beginning

Table 1. SCM and GCM simulations carried out in this study.

Simulation name	Relaxation time	Version
SCM_P5 (3 h)	3 h	Post-CMIP5
SCM_P5 (6 h)	6 h	Post-CMIP5
SCM_P5 (12 h)	12 h	Post-CMIP5
SCM_P5 (24 h)	24 h	Post-CMIP5
SCM_C5 (24 h)	24 h	CMIP5
GCM_P5	–	Post-CMIP5
GCM_C5	–	CMIP5
No Relax	None	Post-CMIP5

of a month, ran the whole month, and were reinitialized every month. Both C5 and P5 simulations from GISS were added in our experiment.

Figure 1 illustrates the model-simulated 500-hPa temperatures with different relaxation time scales against ARM continuous forcing during the first 100 h of November 2007. The simulation without relaxation, as expected, performs the worst in terms of deviations from observations. The simulation with 3-h relaxation agrees well with observations, showing no significant difference in the first 100 h. Simulations with 24-h relaxation have more deviations from observations than the 3-h simulation, due to the lesser constraint placed by the forcing. The mean error from different SCM runs during the entire month is consistent with the results seen in Fig. 1 (not shown).

3. Results and discussion

3.1. *Atmospheric state*

Given the central role of RH in GCMs as a regulator of longwave radiation and an indicator of when clouds will form, it is necessary to assess simulated RH together with temperature against observations. Figure 2 shows the mean biases of simulated temperature and RH profiles using three different relaxation time scales against ARM forcing data during the period 2002–08. As shown in Fig. 2a, simulated temperatures had positive biases above 200 hPa and below 850 hPa, but had negative biases from 850 hPa to 200 hPa. The SCM_P5 (24 h) and SCM_C5 (24 h) simulations had the largest temperature bias (1 K) at all levels, reaching up to 2 K near the surface. Biases for the simulation using a 3-h relaxation time were generally within 0.2 K and up to 1 K near the surface. The shorter relaxation time more strongly constrains temperature, making the simulations agree better with observations.

For the simulated RH (Fig. 2b), the pattern in biases is generally opposite to that seen for temperature. The opposite signs in temperature and RH biases make physical sense because higher (lower) temperatures tend to have lower (higher) RH for a given specific humidity. The RH bias is within 5% at all levels when using the 3-h relaxation time. Simulations with 24-h relaxation have a more positive bias (~10%), especially around 800 hPa. Because the SCM_P5 (24 h) simulation tended to produce more moisture in the atmosphere than the SCM_P5 (3 h) simulation, more clouds may have been generated because RH plays an essential role in simulating stratiform clouds in the model. Note that the continuous forcing used as ground truth is not error free. The ARM forcing has a nearly 5% positive RH bias near the surface compared to Cloud Modeling Best Estimate soundings, as discussed by Kennedy et al. (2011). Therefore, the large negative bias seen in the model simulations can be partly ascribed to the forcing bias.

3.2. *CF*

3.2.1. *Vertical distribution*

Figure 3 shows the monthly mean time–height series of CFs observed by the ARM MMCR (Fig. 3a), simulated by SCM_P5 using different relaxation times (Figs. 3b and c), SCM_C5 (Fig. 3d), and P5 and C5 GCM output (Figs. 3e and f) from a grid box over the ARM SGP site during the period 2002–08. The GCM results will be discussed in section 3.4. Figure 3a shows a bimodal distribution of clouds with peaks around 850 hPa and 300 hPa from October to June. High clouds were present nearly all year round, reaching a maximum in February and June. Low clouds persisted from October to June and occurred the least during the summer months.

Compared to the ARM MMCR-derived CF, all SCM simulations can capture the seasonal variation in and vertical distribution of CFs, but each simulation has its own characteristics due to different relaxation time scales or model versions. The SCM_P5 (3 h) run simulated the smallest CFs among all simulations, especially for high-level clouds, regardless

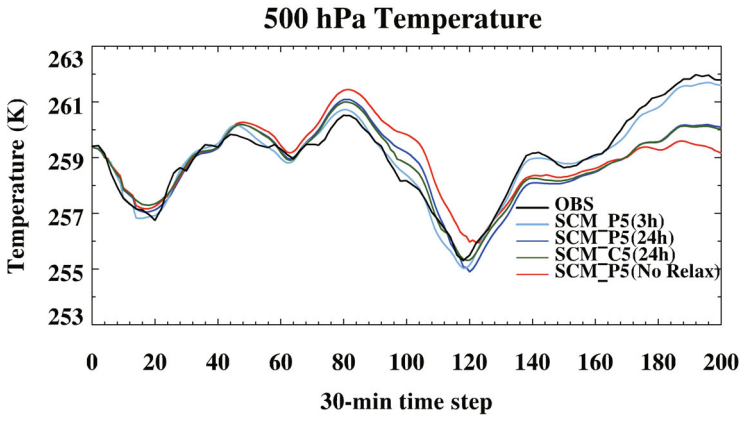

Fig. 1. 500-hPa temperature fields provided by ARM continuous forcing (black line) and SCM simulations during the first 100 h of November 2007. Each step is 30 min in the SCM model. SCM_C5 and SCM_P5 represent the GISS CMIP5 and post-CMIP5 SCMs, respectively. Simulations with different relaxation times (3 or 24 h) are used on both temperature and humidity fields.

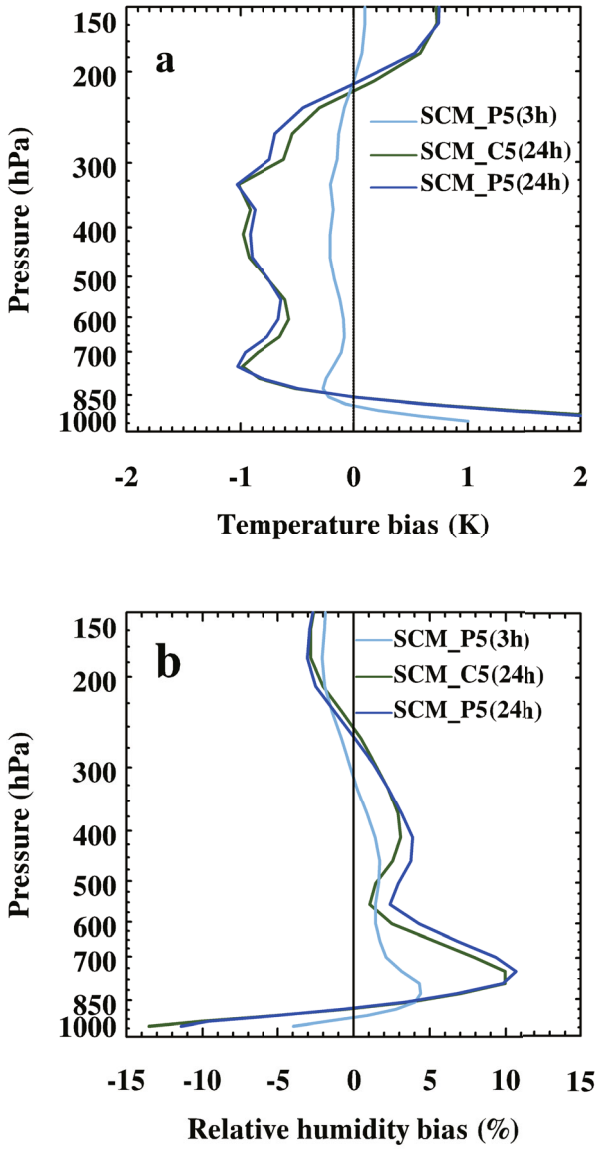

Fig. 2. Vertical profiles of mean biases of simulated (a) temperature and (b) humidity with respect to ARM forcing data during the period 2002–08.

of season. CFs were underestimated at all levels under 200 hPa. Compared to the SCM_P5 (3 h) run, more high clouds were simulated with the longer relaxation time (24 h), which agreed well with ARM MMCR observations, but low-level CFs were less than those from the SCM_P5 (3 h) run, especially in the early spring. Differences in CFs arising from using different relaxation times occurred because the model tended to produce a moist bias, which created a moist atmosphere under a weak constraint, thus producing more high and mid-level clouds and fewer low clouds. The cloud simulations are also consistent with the RH bias seen in Fig. 2b.

For the C5 version SCM (24 h), the general pattern in cloud vertical distribution is similar to the SCM_P5 (24 h) version. Both produce more high and mid-level clouds than SCM_P5 (3 h). The largest difference seen in the P5 simulations was in the lower level of the atmosphere, particularly around 850 hPa, where C5-simulated low clouds were persistently much higher (lower) than the P5 results above (below) 850 hPa. Unlike observations and the P5 version, no obvious seasonal variation was found in the C5 version. The vertical distribution of the SCM_C5 simulated clouds was discontinuous and cut off around 850 hPa, which likely happened because of the poor PBL parameterization in the C5 version.

To quantitatively estimate the CF differences between observations and simulations, the vertical distributions of observed and simulated CFs were averaged over the seven-year period (Fig. 4). From the discussion about Fig. 3 and as shown in Fig. 4, both observed and simulated CFs had bimodal distributions with weaker peaks around 850 hPa (~12.2% from the ARM MMCR) and strong peaks at 300 hPa (~15.5% from the ARM MMCR). Both SCM_P5 simulations captured this vertical structure of CFs with some discrepancies at different levels. Above 200 hPa, the CF differences between MMCR retrievals and SCM_P5 simulations were almost negligible. The largest discrepancy occurred from 300–700 hPa, and especially around 300 hPa, where the SCM_P5 (3 h)-simulated CFs were about 5% lower than MMCR observations and the SCM_P5 (24 h) CFs were nearly identical to MMCR-observed CFs. This result is consistent with the RH comparisons, where SCM_P5 (24 h) simulations had the largest positive RH biases in the middle and high levels of the atmosphere (Fig. 2c). Below 700 hPa, SCM_P5-simulated CFs followed the observed pattern in vertical distribution, although the magnitudes of the CFs were about 3% lower than the MMCR CFs. Near the surface, the SCM_P5 (24 h)-simulated CFs were approximately 5% lower than the MMCR CFs due to its 10% negative bias in RH. The SCM_C5 (24 h) model performance was almost identical to that of the P5 model above 500 hPa. However, the C5 model version was unable to capture the pattern in the middle and low levels of the atmosphere. Too much cloud was produced above 850 hPa and too little was simulated below 850 hPa. Twin peaks in vertical distribution were seen in the C5 simulation and not in the observations or the P5 simulations. These different vertical distributions in low cloud cover reflect the differences in the turbulence parameterization, which is most likely related to the different methods used for calculating the PBL height in the P5 and C5 versions. Figure 4 also illustrates that the convective clouds simulated by SCMs are negligible, with CFs less than 1% at all levels, suggesting that stratiform clouds are dominant over the ARM SGP site.

It is noteworthy that CF retrievals can differ when using different remote sensing instruments. In this study, MMCR-derived CFs were used as ground truth data to evaluate model simulations. The MMCR can observe the majority of clouds within the atmospheric column, but experiences strong attenuation during heavy precipitation events and has difficulty in detecting optically thin cirrus clouds. The MPL compliments the MMCR well by detecting optically thin cirrus clouds, but its signals can be attenuated quickly when optically thick clouds are present. The mean CF increased by 10% when using a combination of MMCR and MPL measurements during the seven-year period, mainly through an increase in optically

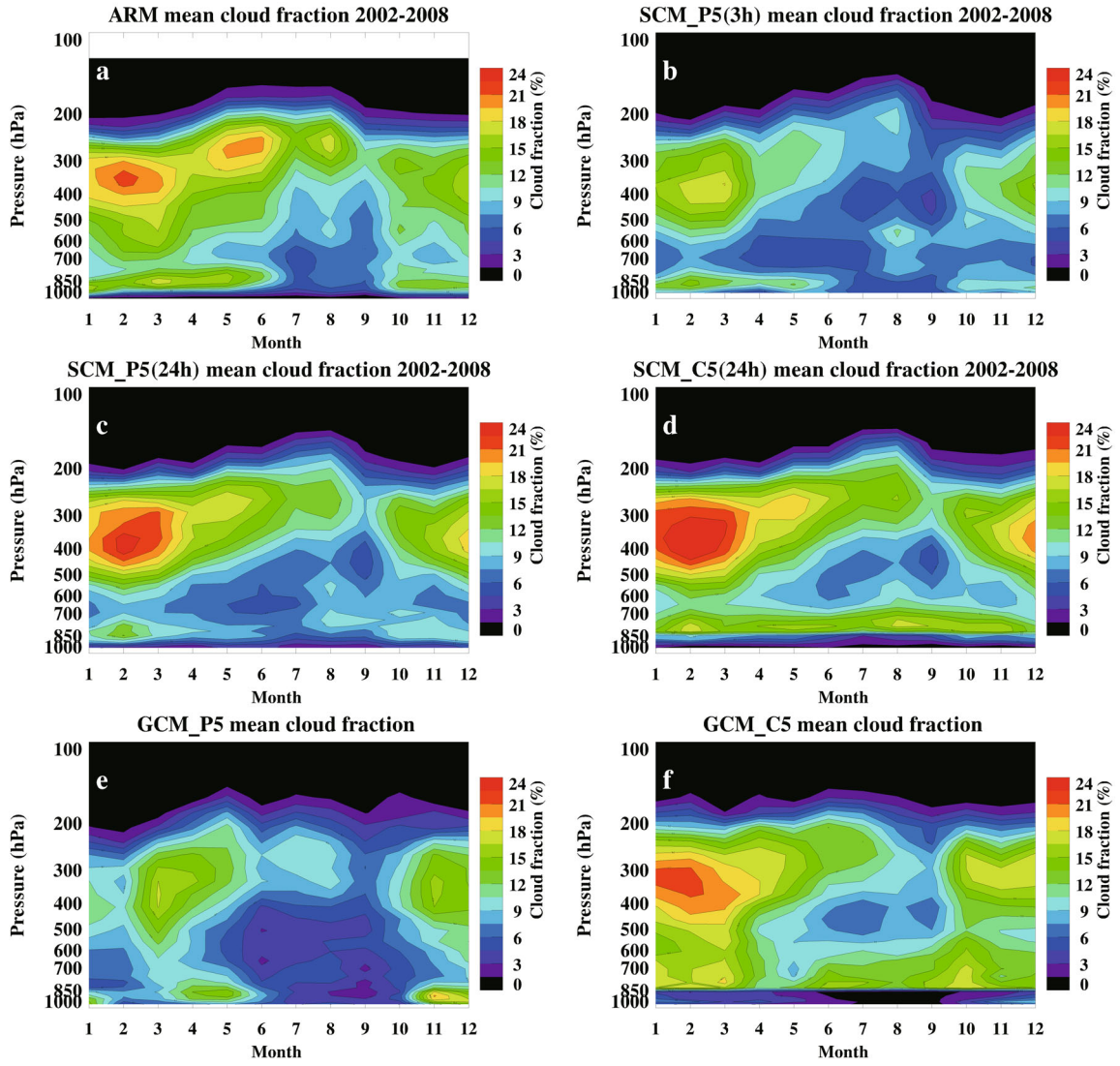

Fig. 3. Monthly mean time–height CFs derived from (a) ARM SGP ground-based MMCR observations, and NASA GISS post-CMIP5 [except CMIP5 in (d) and (f)] simulations with a grid box of 2° (lat) ×2.5° (lon) over the ARM SGP site during the period 2002–08, (b) SCM_P5 simulations with 3-h relaxation on both temperature and humidity fields, (c) SCM_P5 simulations with 24-h relaxation on both temperature and humidity fields, (d) SCM_C5 simulations with 24-h relaxation on both temperature and humidity fields, (e) output from the GISS post-CMIP5 GCM simulation over the ARM SGP site, and (f) output from the GISS CMIP5 GCM

thin high clouds that are present year round and mid-level clouds that are present during the summer months. However, MMCR-derived CFs were still used as ground truth in this study because the models had difficulty in simulating optically thin clouds and passive satellites could not detect them. Therefore, it is meaningful to compare the MMCR-derived and model-simulated CFs.

3.2.2. *Monthly mean*

Figure 5 shows the monthly mean total and high/mid/low CFs derived from ARM MMCR observations and model simulations at the ARM SGP site. Their annual averages are listed in Table 2. The ARM MMCR-derived total CFs remained high (~50%) during winter and spring months, reached a minimum (~40%) during July–September, and then increased during October–December, which agrees well with other studies (Dong et al., 2006; Xi et al., 2010). Monthly uncertainties of total CF observed by the MMCR instrument were calculated using a bootstrap technique (Kennedy et al., 2013) for months with instrument uptimes of >95%. The MMCR-observed total CFs are robust, with small uncertainties (2%), for most months. The uncertainty for MMCR CFs during 2002–03 is relatively larger due to more frequent downtime of the instrument (not shown). Compared to the seasonal variation in the ARM MMCR total CF (Fig. 5a), all SCM-simulated CFs had similar seasonal changes with relatively large differences in some months, except for the SCM_C5 (24 h) simulation. In general, the SCM_P5 (24 h)-simulated CFs were closer to the MMCR observations (45%), with annual differences within ±1%. On the other

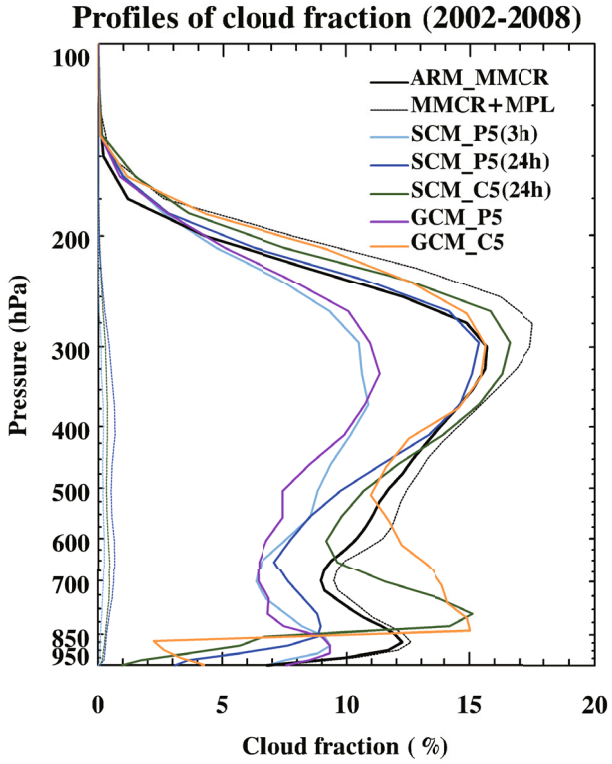

Fig. 4. Mean vertical distributions of total CF derived from ARM SGP MMCR and MPL observations, and simulated by different SCM and GCM runs for the period 2002–08. The dashed non-black lines represent CFs produced by the GISS convection scheme. The dashed black line represents the CF derived from ARM MMCR-MPL measurements.

hand, the SCM_P5 (3 h)-simulated CFs were much less than the ARM MMCR CFs throughout the year, with a 7% negative bias compared to the MMCR annual total CF. Compared to the P5 simulations, the SCM_C5-simulated CFs were generally higher all year round, particularly during the period of June–September, which is a strong convection season at the SGP site. The annual averaged SCM_C5 CF was 10% higher than the MMCR annual averaged CF, with a maximum positive bias of 40% during August. Since the same ARM forcing data were used by the SCM_P5 and SCM_C5 models, this large positive bias presumably results from their different PBL parameterizations.

To further investigate the source of biases in total CFs, clouds were classified as high (>6 km), middle (3–6 km), and low (<3 km) clouds. The seasonal variations of these three levels of CFs are presented in Figs. 5b–d and their annual averages are listed in Table 2. The SCM_P5 (3 h) simulation underestimated the high-level CF compared to ground observations by 8%, which resulted in a 7% negative bias in its total CF. After applying a longer relaxation time, the high-level CFs were increased by 7% in the SCM_P5 (24 h) simulation, where the 24-h simulated high-level CFs were almost identical to the MMCR CFs, especially from late spring to early fall. For mid-level clouds, all simulations produced more clouds than observed, ranging from 1% in the SCM_P5 (3 h) simulation to 6% in the SCM_P5 (24 h) simulation. In comparison with observations, the SCM_C5 (24 h) simulation overestimated mid-level CFs by 14%. Most of the overestimation occurred during the summer months, especially in August, which most likely contributed to the RH bias in the forcing. For low-level clouds, the P5 (3 h) simulation performed best (1% positive bias), while other simulations had ±2% biases. Although other simulations agreed within 2% with the MMCR annual result, the good agreement for annual mean was a compensation of negative biases in spring and positive biases during summer.

3.2.3. *Diurnal cycle*

Figure 6 shows the diurnal variation in CFs from surface observations, as well as simulations, during the period 2002–08. The MMCR-derived total CF (Fig. 6a) shows that clouds began to increase in the early morning, reached a maximum at noon, then decreased in the afternoon and dropped to a minimum at 1800 Local Time (LT, Local hour at the ARM SGP is UTC–6 h). The SCM_P5 simulations peaked at 0500 LT and bottomed-out at around 1800 LT. The SCM_P5 (3 h) simulation showed the strongest diurnal variation, with a magnitude of 15%. Similar to its seasonal variation, hourly mean total CFs from the SCM_P5 (3 h) simulation were lower than MMCR observations and other SCM simulations throughout the day. Hourly mean total CFs from the SCM_C5 simulation did not capture the diurnal pattern seen in MMCR observations, showing an opposite diurnal variation in the daytime. The diurnal cycle of CF in the warm season (April to October) showed no significant difference from the whole year, even though the diurnal signal was somewhat enhanced (not shown).

The MMCR-observed high clouds were slightly more frequent during the evening and night, reaching a minimum at ~1400 LT. In contrast to the observed diurnal variation, simulated high-level CFs increased from morning to 1300 LT and then decreased until 2100 LT with more clouds in the daytime than at night. No obvious diurnal variations in mid-level clouds were seen from observations and P5 simulations. The SCM_P5 (3 h) simulation had an excellent agreement with MMCR observations, but both SCM_P5 (24 h) and SCM_C5 simulations overestimated mid-level clouds at all times. The noticeable increase in mid-level CF in the afternoon from the SCM_C5 simulation contributed the most to the total CF from this simulation. MMCR-observed and model-simulated low-level clouds had strong diurnal cycles, which is primarily attributed to solar radiation and the ground temperature. As shown in Fig. 6d, MMCR-observed low-level clouds monotonically increased from 0000 LT to 1000 LT, decreased until 1800 LT, and then leveled off. Simulated P5 low-level clouds peaked at 0700 LT and reached a minimum at 1700 LT. The diurnal variation in low clouds from the C5 simulation agrees well with observations.

3.2.4. *Possible causes of CF bias*

To investigate the possible causes of the biases in simulated CFs, probability distribution functions (PDFs) of high,

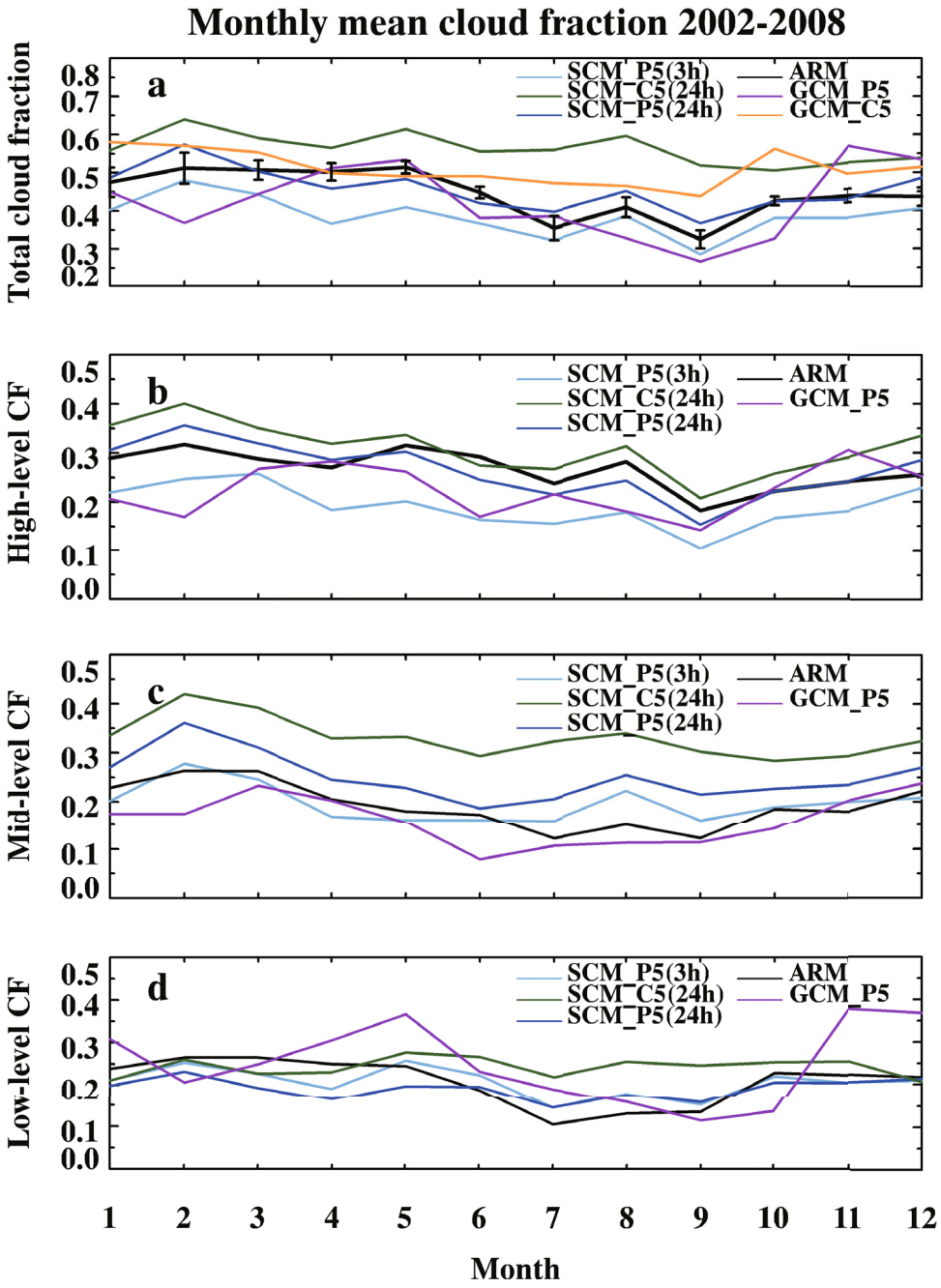

Fig. 5. Monthly mean (a) total CF, (b) high-level CF (>6 km), (c) mid-level CF (3–6 km), and (d) low-level CF (<3 km) derived from ARM MMCR measurements and simulated in a 2° (lat) ×2.5° (lon) grid box centered on the ARM SGP site. The period covered is 2002–08.

Table 2. Mean CF, precipitation rate, and LWP from observations and simulations for the period 2002–08.

	CF (total)	CF (high)	CF (mid)	CF (low)	Precipitation rate (mm d^{-1})	LWP (g m^{-2})
ARM observations	0.45	0.27	0.19	0.21	2.44	80.3
SCM_P5 (3 h)	0.38	0.19	0.20	0.20	1.32	68.8
SCM_P5 (6 h)	0.39	0.19	0.22	0.19	1.48	71.3
SCM_P5 (12 h)	0.41	0.21	0.23	0.19	1.71	74.2
SCM_P5 (24 h)	0.46	0.26	0.25	0.19	2.07	77.2
SCM_C5 (24 h)	0.55	0.30	0.33	0.24	2.26	113.2
GCM_P5	0.42	0.22	0.16	0.25	2.24	77.0
GCM_C5	0.51	–	–	–	2.24	135.8

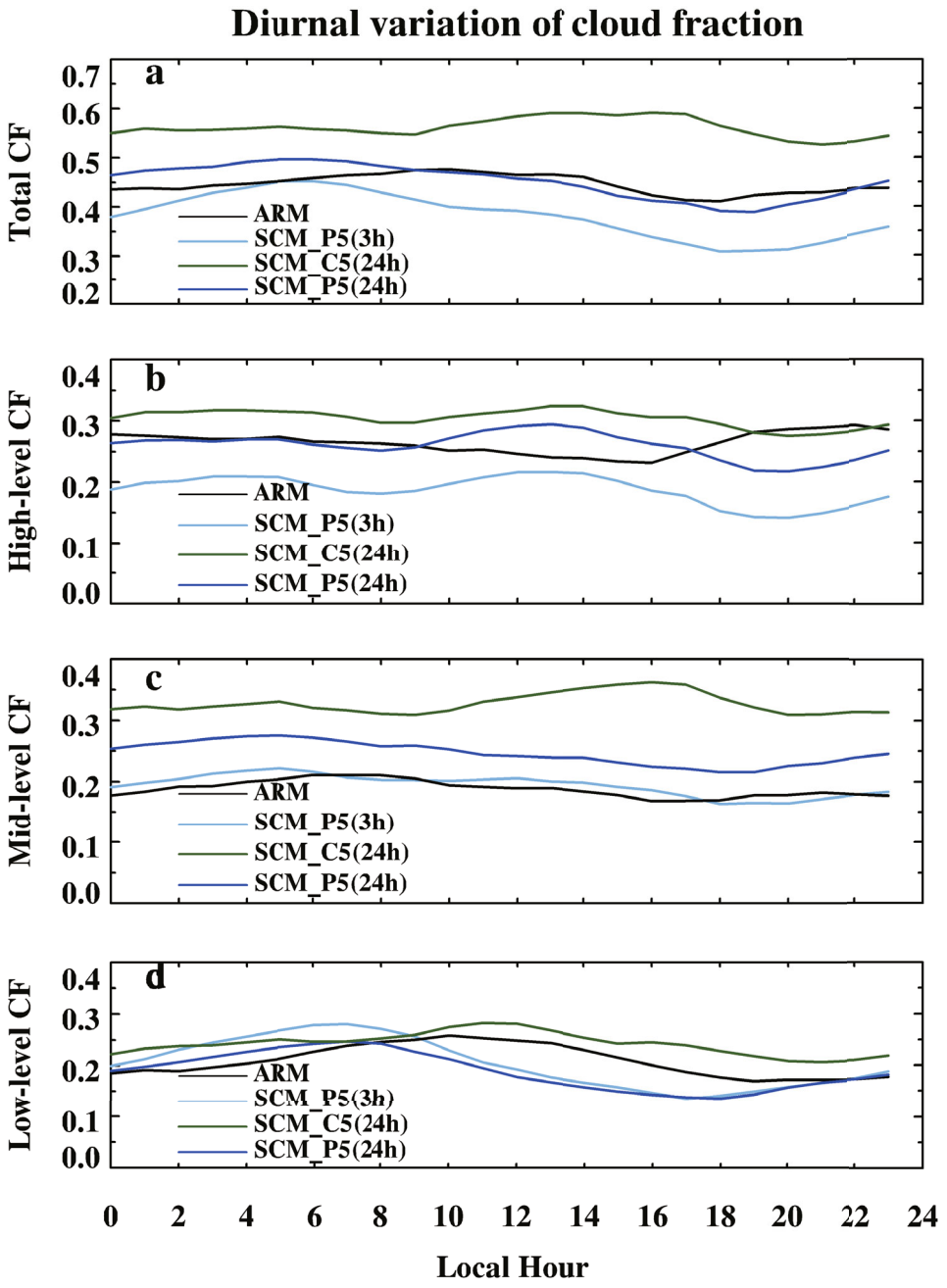

Fig. 6. Diurnal variations in (a) total CF, (b) high-level CF (>6 km), (c) mid-level CF (3–6 km), and (d) low-level CF (<3 km) derived from surface and satellite retrievals and SCM simulations over the ARM SGP site. The period covered is 2002–08. The local hour at the ARM SGP site is calculated as UTC–6 h.

middle and low CFs as a function of RH were calculated (Fig. 7). Whenever the CF is greater than zero, the corresponding RH value is counted. The averaged 350–250 hPa, 550–400 hPa and 950–750 hPa levels for RH were selected to study high-, middle- and low-level clouds. The PDFs for SCMs with different relaxation times are almost identical. Here, P5 and C5 simulations (with a 24-h relaxation time) were compared with observations. The PDF for mid-level clouds from the SCM simulations matched well with observations (Fig. 7b). The largest discrepancies occurred in the high and low levels of the atmosphere. The observed high clouds have a near-Gaussian distribution, while the SCM_P5 simulation has a narrower and steeper distribution, with peaks at around 50% and 60%, respectively (Fig. 7a). The PDF of the SCM_C5 simulation is less steep and thus represents high-level CFs slightly better than the P5 simulation. Simulated high clouds tend to occur under higher RH conditions than observed high clouds, which suggests that the threshold of RH for cloud production in the GISS model might be much larger than what occurs in nature. For low-level clouds (Fig. 7c), SCM_P5-simulated CFs and observations have similar patterns, peaking around RH levels of 70%–80%, while

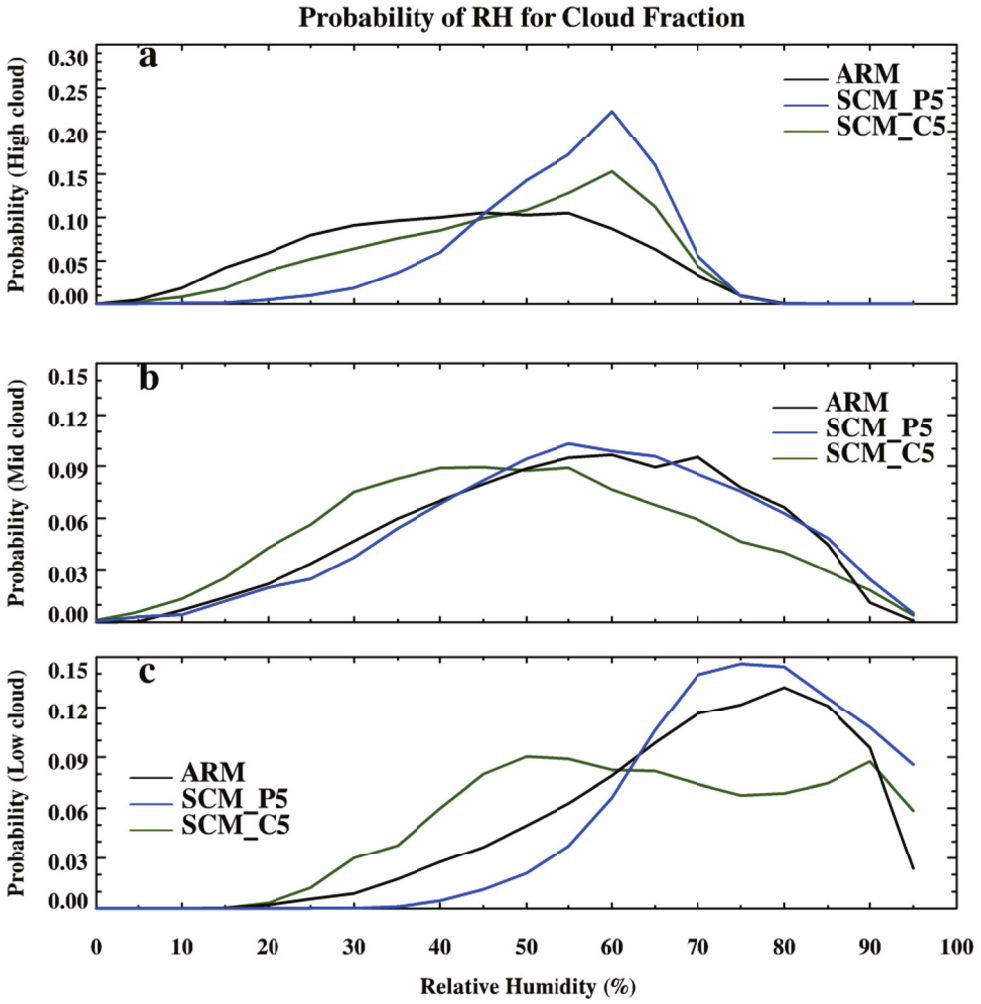

Fig. 7. Probabilities of observed, SCM_P5 (24 h) and C5 simulated (a) high clouds, (b) mid clouds, and (c) low clouds, as a function of RH.

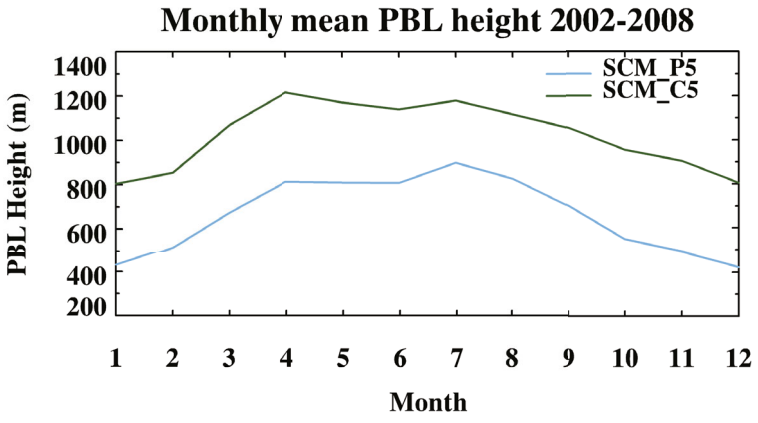

Fig. 8. Monthly mean Post-CMIP5 (blue line) and CMIP5 (green line) SCM-simulated boundary layer heights. The period covered is 2002–08.

SCM_C5-simulated CFs peak around RH = 45% and remain steady above 50%. Thus, the SCM_P5 model appears to simulate low clouds more realistically than the SCM_P5 model. The underestimation in high-level CFs may be primarily due to their weaker relationship with RH and also to a possibly incorrect threshold of RH used in the cloud scheme.

A new scheme is implemented in the P5 model turbulence parameterization, which would contribute to the low-level CF differences seen between the SCM_P5 and SCM_C5 simulations. The new scheme uses the "Richardson number criterion" instead of the "TKE criterion". These two different methods affect the PBL height significantly. Figure 8 shows

the boundary layer heights calculated by the SCM_P5 and SCM_C5 models. The P5-simulated PBL heights are lower than those from the C5 simulations. The differences in PBL height agree well with the differences seen in CF simulations. The higher PBL height tends to transfer moisture higher into the atmosphere, thus leading to a higher CF. Although the shallow convective cloud cover is small compared to the total cloud cover, it could moisten the lower troposphere by vertical transport, and thus stratiform cloud could have a better chance of forming (Del Genio et al., 2012; Yao and Cheng, 2012), in turn resulting in the increase of total CF. Therefore, a positive low-level bias was found in the SCM_C5 simulation.

3.3. *Precipitation and LWP*

Figure 9a shows that there was a strong seasonal variation in observed precipitation at the SGP site during the period 2002–08. Precipitation was greatest in June, when convection occurred frequently over this site. Other local maxima in precipitation occurred in March, August, and October. Precipitation in Oklahoma is primarily controlled by baroclinic wave activity from October to April, and by weaker synoptic systems and mesoscale forcing-produced convection during the warm season (Kennedy, 2011).

GISS SCM-produced precipitation is the sum of convective and stratiform precipitation. All SCM-simulated precipitation mimicked the variation in observed precipitation with different negative biases. Similar to the total CF comparison, the SCM_P5 (24 h)-simulated precipitation (2.07 mm d^{-1}) was closest to observations (2.46 mm d^{-1}), while the SCM_P5 (3 h) result had the greatest difference, with a negative bias of −1.13 mm d^{-1}, especially during the strong convective season from April to September. Similar results were reported by Song et al. (2013), who also used a 3-h relax-

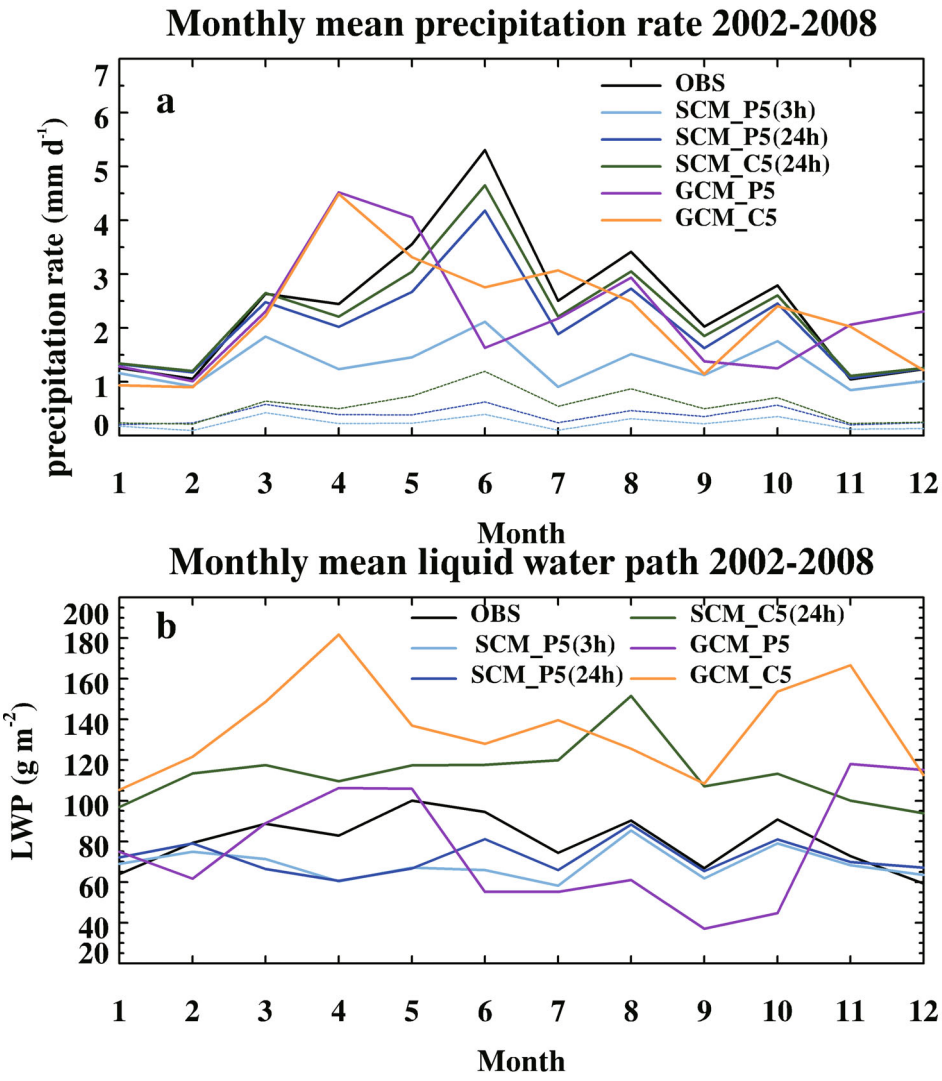

Fig. 9. Monthly mean (a) total (stratiform+convection) precipitation rate measured by ARM surface instruments (black line), and (b) cloud LWP retrieved from ARM microwave radiometer brightness temperature measurements (black line), as well as simulations from different model runs (non-black lines). The period covered is 2002–08. The dashed lines in (a) represent the convective precipitation produced by models.

ation time period for their simulations carried out over the time period of 1999–2001. SCM_C5 simulated precipitation, using a 24-h relaxation time scale, was almost identical to precipitation from the SCM_P5 (24 h) simulation. All these results suggest that a 3-h relaxation may constrain the SCM too closely, which could inhibit the generation of clouds and precipitation from the model. In contrast to SCM_P5 (3 h) results, the precipitation rate when using a longer relaxation time (24 h) increased significantly and was close to observations due to accumulated moisture (Fig. 2). The SCM_P5 (24 h) results are also similar to results from the SCM simulation conducted by Kennedy (2011), who reset the simulation every 24 h.

The convective parts of simulated precipitation are also shown in Fig. 9a, using dashed color lines. Precipitation produced by convection plays a minor role in producing total precipitation. However, during the convective season, the majority of precipitation in models is expected to be convective, as shown by satellite observations. This weak convection related to the parcel-lifting-based trigger used in the convection scheme is always an issue in the GISS model. The turbulent kinetic energy is not strong enough to provide updrafts capable of lifting air parcels to the level of free convection. Such an issue indicates that the convection scheme in the GISS model still needs to be improved.

The diurnal cycle of the precipitation rate is shown in Fig. 10a. The observed diurnal cycle varied sinusoidally with a local maximum at 0300 LT and minimum at 1300 LT. Although all simulations captured the observed diurnal variation, their variations were less distinct than observations. Similar to the seasonal comparison, the simulations with a 24-h relaxation time were closest to observations, while 3-h simulations were much lower than observations. In the warm season (March–September), the diurnal signal increased for both observations and simulations (not shown), but the diurnal pattern remained consistent with results for the whole year. The diurnal cycles in the warm season are more obvious, but the general pattern is consistent (not shown).

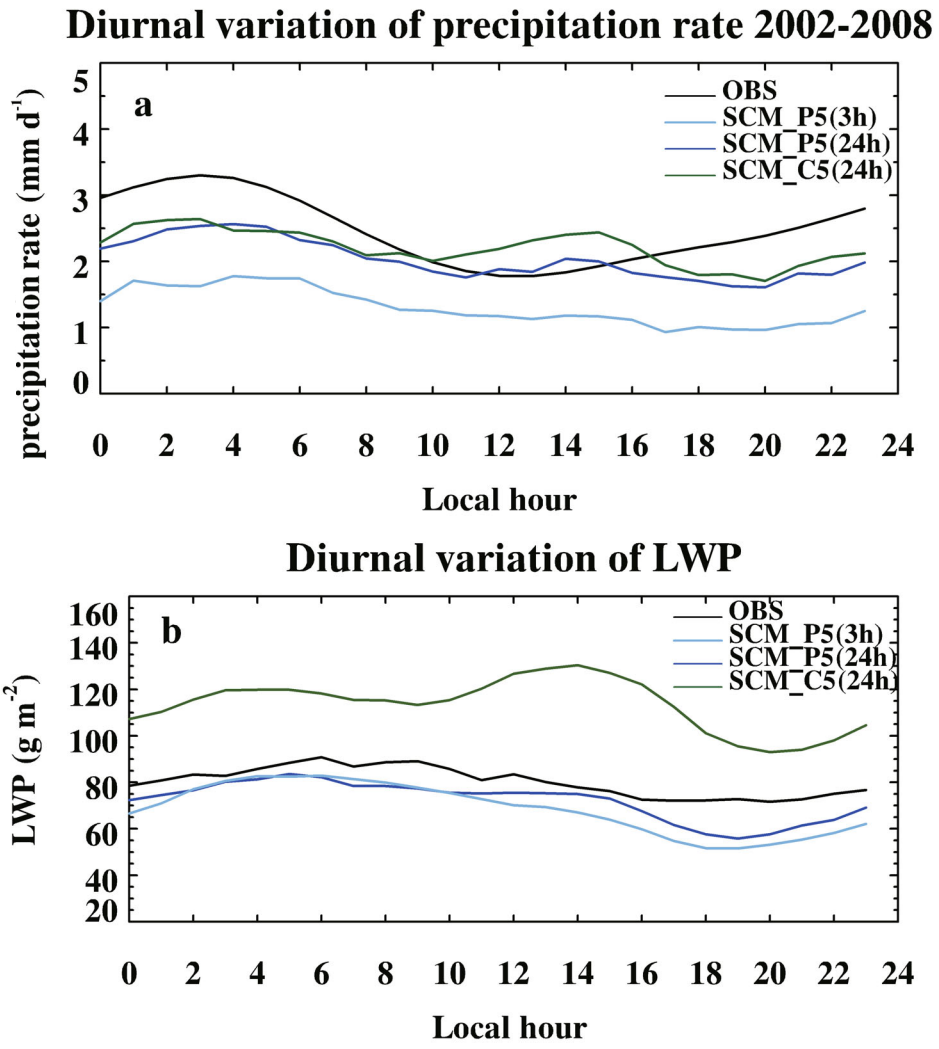

Fig. 10. Diurnal variations in (a) total (stratiform+convection) precipitation rate measured by ARM surface instruments (black line), and (b) cloud LWP retrieved from ARM microwave radiometer brightness temperature measurements (black line), as well as simulations from different SCM runs (non-black lines). The period covered is 2002–08.

To further investigate the seasonal and diurnal variations in precipitation, monthly mean and diurnal variations in LWP were examined (Figs. 9b and 10b). Figure 9b shows monthly mean LWP retrieved from the ARM microwave radiometer and simulated in different model runs during the period 2002–08. The seasonal variation in observed LWP was similar to observed precipitation, with a peak during May–June and local maxima in August and October. The P5-simulated LWPs were almost identical to the observed LWPs from July to December, but were underestimated from March to June. The C5 model captured the seasonal variations but overestimated LWP every month. The diurnal variation in LWP is shown in Fig. 10b, where both observed and simulated LWPs have sinusoidal-like variations with maxima around 0600 LT and minima around 1900 LT. Both monthly and hourly mean SCM_C5-simulated LWPs were higher than observed and P5-simulated LWPs. The C5 model had some issues in simulating the diurnal pattern.

Based on the discussion about clouds and precipitation, and the annual averages listed in Table 2, we conclude that the SCM_P5-simulated CFs and LWPs increase when relaxation time scales change from 3 to 24 h. For instance, the CF increased from 0.38 in the SCM_P5 (3 h) simulation to 0.46 in the SCM_P5 (24 h) simulation, which is closest to observations. The LWP increased from 68.8 g m^{-2} in the SCM_P5 (3 h) simulation to 77.2 g m^{-2} in the SCM_P5 (24 h) simulation, which is also closest to observations. The SCM_C5-simulated CFs, precipitation, and LWPs are all higher than the P5 simulations with a 24-h relaxation time scale. The P5-simulated precipitation increased from 1.32 mm d^{-1} in the SCM_P5 (3 h) simulation to 2.07 mm d^{-1} in the SCM_P5 (24 h) simulation. Therefore, we conclude that the simulated CF and LWP are primarily determined by the cloud parameterizations in different model versions, although they also increase with longer relaxation time scales. However, simulated precipitation depends heavily on the relaxation time scale, i.e., precipitation increases significantly with relaxation time scale.

3.4. *GCM and SCM comparison*

In this section we focus on comparing P5 and C5 GCM and SCM simulated clouds and precipitation using ARM ground-based observations as ground truth. Since both the GCM and SCM use the same physical parameterizations, it is expected that any differences seen arise from the different dynamics implemented in the models. As shown in Fig. 3d and Fig. 4, GCM simulations are similar to their corresponding SCM simulations, i.e., the same version of the GCM and SCM generate similar CFs. The CF profiles generated by the P5 GCM were much lower than observations at all levels, consistent with the simulated CF from the SCM_P5 (3 h) model. For the C5 GCM, the same issues concerning CF within the PBL arose and more biases were seen in the mid-level part of the atmosphere.

Although the annual mean P5 GCM-simulated total CF was only 3% less than the MMCR-retrieved CF (Table 2), seasonal variations in the monthly mean did not follow observations. P5-simulated total CFs varied from 0.267 in September to 0.571 in November. Of all the simulations, the C5 GCM-simulated total CF showed no significant seasonal changes. The GCM_P5-simulated mean high-level CF was 5% less than the MMCR CF, falling somewhere between the SCM simulations. The GCM_P5-simulated mean mid-level CF was 3% less than the MMCR mid-level CF, which is much less than other SCM simulations. However, the mean low-level CF was 4% higher than the MMCR low-level CF and had a ~15% positive bias from November to December. Note that positive biases in all SCM-simulated low and middle clouds in August did not show up in the GCM simulations. This is attributed to the dynamic difference between the GCM and the SCM driven by the ARM forcing.

For GCM-simulated precipitation rates, the C5 and P5 GCMs have the same annual mean (2.24 mm d^{-1}), with slight differences seen in certain months. Although their annual averages were close to the observed average, their monthly means showed significant patterns compared with observations and SCM simulations. GCM-simulated precipitation was out of phase. For example, a peak is seen in April, which is two months earlier than the observed and P5-simulated peaks. This is probably related to the model's inability to capture the convection propagating from the Rockies, which is a common issue in many models. The similar annual averages among the GCM and SCM_P5 simulations and observations, and the complicated out-of-phase seasonal variations, suggest that the same physical cloud parameterization can generate similar statistical results over a long time period, but different dynamics can drive the differences in seasonal variations.

4. Summary and conclusions

The latest NASA GISS-E2 SCM simulations of CF, LWP, and precipitation were evaluated against ARM SGP ground-based observations for the study period 2002–08. In addition to investigating the responses of the post-CMIP5 (SCM_P5) simulations to various relaxation time scales, we compared SCM_P5 results with CMIP5 (SCM_C5) simulations using ground-based observations as ground truth to explore the impact of their physical parameterizations on simulated results. GCM results were used to examine whether the responses of GCM and SCM simulations to the improved parameterization were consistent. The key findings are summarized as follows:

(1) Compared with MMCR retrievals at the SGP site, the CMIP5 SCM overestimated the total CF (10%). The positive bias arises from the overestimation of mid-level clouds (33% versus 19% from observations) and low-level clouds (24% versus 21% from observations). The vertical profile of CF is unrealistic within the boundary layer, producing too many (few) clouds above (under) 850 hPa. The low-level structure of and seasonal variation in CF simulated by the latest post-CMIP5 SCM was improved due to the implementation of a new turbulence parameterization scheme. However, the underestimation of high cloud related to the weaker relationship between CF and RH suggests that the cloud scheme needs more development.

(2) Cloud and precipitation simulations from the SCM are sensitive to the relaxation time scale. With a longer relaxation time scale (24 h), clouds and precipitation are in better agreement with ground observations, although state variables develop more biases. The post-CMIP5 SCM-simulated total CF increased from 38% to 46%, mostly due to an increase in high-level clouds. Simulated precipitation increased from 1.32 mm d^{-1} to 2.07 mm d^{-1}, which was closer to the observed value (2.24 mm d^{-1}). The less accurate temperature and RH fields leading to the more accurate representation of clouds suggests a deficiency in the model physics.

(3) GCM and SCM simulations from the same version agree well and have consistent responses to the modified turbulence parameterization scheme. Both the CMIP5 SCM and GCM have difficulty in simulating the vertical structure of and seasonal variation in low-level clouds. This situation is greatly improved in the post-CMIP5 GCM and SCM. In addition, since both the GCM and SCM used the same physical parameterizations, differences in their respective simulated results are due to the dynamic scheme used in the GCM, because we assume that the ARM forcing driving the SCM properly represents the dynamics. Our study shows that although the annual averages of the GCM-simulated total CF, LWP, and precipitation were close to observations and SCM simulations, their monthly means were out of phase and unable to capture the seasonal variation. The similar annual averages among the GCM and SCM_P5 simulations and observations, and the complicated out-of-phase seasonal variations, suggest that the same physical cloud parameterization can generate similar statistical results over a long time period, but different dynamics can drive the differences in seasonal variations.

Overall, the changes implemented in the latest GISS GCM, especially the changes in boundary layer height, have significantly improved simulated clouds. The change of PBL stability affects the shallow convection, and thus moistens the atmosphere and contributes to the formation of stratiform clouds. The SCM is a useful tool to test and investigate physical parameterizations, but due to its lack of large-scale dynamics, the SCM-simulated seasonal cycle is probably inconsistent with that simulated by the GCM, as shown in this study. While the results of this study suggest some possible directions for future improvement, simulations and analyses need to be carried out on data from other climate regimes, such as tropical ocean regions, if forcing data are available, so that a general and more comprehensive understanding of the possible deficiencies in the model can be gained.

Acknowledgements. Data were obtained from the ARM program sponsored by the U.S. Department of Energy Office of Energy Research, Office of Health and Environmental Research, Environmental Sciences Division. The data can be downloaded from http://www.archive.arm.gov/. Researchers at the University of North Dakota were supported by the DOE ASR program (Grant No. DE-SC008468). Special thanks are extended to Mr. Ryan Stanfield, who provided the NASA GISS post-CMIP5 output over the ARM SGP region.

REFERENCES

Ackerman, T. P., and G. M. Stokes, 2003: The atmospheric radiation measurement program. *Physics Today*, **56**, 38–44.

Benjamin, S. G., and Coauthors, 2004: An hourly assimilation/forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518.

Betts, A. K., and C. Jakob, 2002: Study of diurnal cycle of convective precipitation over Amazonia using a single column model. *J. Geophys. Res. Atmos.*, **107**, ACL 25-1–ACL 25-13.

Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchand, M. A. Miller, and B. E. Martner, 2000: Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *J. Appl. Meteor.*, **39**, 645–665.

Del Genio, A. D., and M. S. Yao, 1993: Efficient cumulus parameterization for long-term climate studies: The GISS scheme. *The Representation of Cumulus Convection in Numerical Models*, K. A. Emanuel and D. J. Raymond, Eds., American Meteorological Society, 181–184.

Del Genio, A. D., A. B. Wolf, and M.-S. Yao, 2005: Evaluation of regional cloud feedbacks using single-column models. *J. Geophys. Res.*, **110**, D15S13, doi: 10.1029/2004JD005011.

Del Genio, A. D., Y.-H. Chen, D. Kim, and M.-S. Yao, 2012: The MJO transition from shallow to deep convection in *CloudSat*/CALIPSO data and GISS GCM simulations. *J. Climate*, **25**, 3755–3770, doi: 10.1175/JCLI-D-11-00384.1.

Dolinar, E. K., X. Q. Dong, B. K. Xi, J. H. Jiang, and H. Su, 2015: Evaluation of CMIP5 simulated clouds and TOA radiation budgets using NASA satellite observations. *Climate Dyn.*, **44**, 2229–2247.

Dong, X. Q., B. K. Xi, and P. Minnis, 2006: A climatology of midlatitude continental clouds from the ARM SGP Central Facility. Part II: Cloud fraction and surface radiative forcing. *J. Climate*, **19**, 1765–1783.

Ghan, S., and Coauthors, 2000: A comparison of single column model simulations of summertime midlatitude continental convection. *J. Geophys. Res. Atmos.*, **105**, 2091–2124, doi: 10.1029/1999JD900971.

Hack, J. J., and J. A. Pedretti, 2000: Assessment of solution uncertainties in single-column modeling frameworks. *J. Climate*, **13**, 352–365.

Iacobellis, S. F., R. C. J. Somerville, and D. E. Lane, 2000: Analysis of forcing methods for single-column models. *11th Symposium on Global Change Studies*, La Jolla, CA.

Jiang, J. H., and Coauthors, 2012: Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-train" satellite observations. *J. Geophys. Res. Atmos.*, **117**, D14105, doi: 10.1029/2011JD017237.

Jiang, J. H., and Coauthors, 2015: Evaluating the diurnal cycle of upper-tropospheric ice clouds in climate models using SMILES observations. *J. Atmos. Sci.*, **72**, 1022–1044, doi: 10.1175/JAS-D-14-0124.1.

Kennedy, A. D., 2011: Evaluation of a single column model at the Southern Great Plains Climate Research Facility. Ph. D. dissertation, University of North Dakota, 149 pp.

Kennedy, A. D., X. Q. Dong, B. K. Xi, P. Minnis, A. D. Del Genio, A. B. Wolf, and M. M. Khaiyer, 2010: Evaluation of the NASA GISS single-column model simulated clouds using combined surface and satellite observations. *J. Climate*, **23**, 5175–5192, doi: 10.1175/2010JCLI3353.1.

Kennedy, A. D., X. Q. Dong, B. K. Xi, S. C. Xie, Y. Y. Zhang, and J. Y. Chen, 2011: A comparison of MERRA and NARR re-

analyses with the DOE ARM SGP data. *J. Climate*, **24**, 4541–4557.

Kennedy, A. D, X. Q. Dong, and B. K. Xi, 2013: Cloud Fraction at the ARM SGP Site: Instrument and Sampling Considerations from 14 years of ARSCL. *Theor. Appl. Climatol*, doi: 10.1007/s00704-013-0853-9.

Klein, S. A., and A. D. Del Genio, 2006: ARM's support for GCM improvement: A white paper. UCRL-MI-224010, ARM-06-012.

Klein, S. A., Y. Y. Zhang, M. D. Zelinka, R. Pincus, J. Boyle, and P. J. Gleckler, 2013: Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator. *J. Geophys. Res. Atmos.*, **118**, 1329–1342.

Lauer, A., and K. Hamilton, 2013: Simulating clouds with global climate models: A comparison of CMIP5 results with CMIP3 and satellite data. *J. Climate*, **26**, 3823–3845, doi: 10.1175/JCLI-D-12-00451.1.

Li, J.-L. F., D. E. Waliser, G. Stephens, S. Lee, T. L'Ecuyer, S. Kato, N. Loeb, and H.-Y. Ma, 2013: Characterizing and understanding radiation budget biases in CMIP3/CMIP5 GCMs, contemporary GCM, and reanalysis. *J. Geophys. Res. Atmos.*, **118**, 8166–8184, doi: 10.1002/jgrd.50378.

Lohmann, U., N. McFarlane, L. Levkov, K. Abdella, and F. Albers, 1999: Comparing different cloud schemes of a single column model by using mesoscale forcing and nudging technique. *J. Climate*, **12**, 438–461.

Moran, K. P., B. E. Martner, M. J. Post, R. A. Kropfli, D. C. Welsh, and K. B. Widener, 1998: An unattended cloud-profiling radar for use in climate research. *Bull. Amer. Meteor. Soc.*, **79**, 443–455.

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. *Atmospheric Chemistry and Physics*, **12**, 1785–1810.

Randall, D. A., K. M. Xu, R. J. C. Somerville, and S. Iacobellis, 1996: Single-column models and cloud ensemble models as links between observations and climate models. *J. Climate*, **9**, 1683–1697.

Randall, D. A., and D. G. Cripe, 1999: Alternative methods for specification of observed forcing in single-column models and cloud system models. *J. Geophys. Res. Atmos.*, **104**, 24 527–24 545.

Schmidt, G. A., and Coauthors, 2014: Configuration and assessment of the GISS Model E2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, **6**, 141–184.

Song, H., W. Y. Lin, Y. L. Lin, A. B. Wolf, R. Neggers, L. J. Donner, A. D. Del Genio, and Y. G. Liu, 2013: Evaluation of precipitation simulated by seven SCMs against the ARM observations at the SGP site. *J. Climate*, **26**, 5467–5492, doi: 10.1175/JCLI-D-12-00263.1.

Song, H., and Coauthors, 2014: Evaluation of cloud fraction simulated by seven SCMs against the ARM observations at the SGP site. *J. Climate*, **27**, 6698–6719.

Stanfield, R. E., X. Q. Dong, B. K. Xi, A. Kennedy, A. D. Del Genio, P. Minnis, and J. H. Jiang, 2014: Assessment of NASA GISS CMIP5 and post-CMIP5 simulated clouds and TOA radiation budgets using satellite observations: Part I: Cloud fraction and properties. *J. Climate*, **27**, 4189–4208, doi: 10.1175/JCLI-D-13-00558.1.

Stanfield, R. E., X. Q. Dong, B. K. Xi, A. D. Gel Genio, P. Minnis, D. Doelling, and N. Loeb, 2015: Assessment of NASA GISS CMIP5 and post-CMIP5 simulated clouds and TOA radiation budgets using satellite observations. Part II: TOA radiation budget and CREs. *J. Climate*, **28**, 1842–1864, doi: 10.1175/JCLI-D-14-00249.1.

Su, H., and Coauthors, 2013: Diagnosis of regime-dependent cloud simulation errors in CMIP5 models using "A-Train" satellite observations and reanalysis data. *J. Geophys. Res. Atmos.*, **118**, 2762–2780, doi: 10.1029/2012JD018575.

Sundqvist, H., 1978: A parameterization scheme for non-convective condensation including prediction of cloud water content. *Quart. J. Roy. Meteor. Soc.*, **104**, 677–690.

Sundqvist, H., E. Berge, and J. E. Kristjánsson, 1989: Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Mon. Wea. Rev.*, **117**, 1641–1657.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.*, **106**(D7), 7183–7192, doi: 10.1029/2000JD900719.

Wang, H. L., and W. Y. 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, doi: 10.1029/2012JD018619.

Xi, B. K., X. Q. Dong, P. Minnis, and M. M. Khaiyer, 2010: A 10 year climatology of cloud fraction and vertical distribution derived from both surface and GOES observations over the DOE ARM SPG site. *J. Geophys. Res. Atmos.*, **115**, D12124, doi: 10.1029/2009JD012800.

Xie, S. C., R. T. Cederwall, and M. H. Zhang, 2004: Developing long-term single-column model/cloud system-resolving model forcing data using numerical weather prediction products constrained by surface and top of the atmosphere observations. *J. Geophys. Res. Atmos.*, **109**, D01104, doi: 10.1029/2003JD004045.

Xie, S. C., and Coauthors, 2005: Simulations of midlatitude frontal clouds by single-column and cloud-resolving models during the Atmospheric Radiation Measurement March 2000 cloud intensive operational period. *J. Geophys. Res. Atmos.*, **110**, D15S03, doi: 10.1029/2004JD005119.

Xu, K. M., and Coauthors, 2005: Modeling springtime shallow frontal clouds with cloud-resolving and single-column models. *J. Geophys. Res. Atmos.*, **110**, D15S04 doi: 10.1029/2004JD005153.

Yao, M. S., and Y. Cheng, 2012: Cloud simulations in response to turbulence parameterizations in the GISS Model E GCM. *J. Climate*, **25**, 4963–4974, doi: 10.1175/JCLI-D-11-00399.

Yoo, H., and Z. Q. Li, 2012: Evaluation of cloud properties in the NOAA/NCEP global forecast system using multiple satellite products. *Climate Dyn.*, **39**, 2769–2787.

Yoo, H., Z. Q. Li, Y.-T. Hou, S. Lord, F. Z. Weng, and H. W. Barker, 2013: Diagnosis and testing of low-level cloud parameterizations for the NCEP/GFS model using satellite and ground-based measurements. *Climate Dyn.*, **41**, 1595–1613.