**Atmospheric
Chemistry
and Physics**

Evaluation of cloud fraction and its radiative effect simulated by IPCC AR4 global models against ARM surface observations

Y. Qian[1], C. N. Long[1], H. Wang[1], J. M. Comstock[1], S. A. McFarlane[1], and S. Xie[2]

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA
[2]Lawrence Livermore National Laboratory, Livermore, CA, USA

Correspondence to: Y. Qian (yun.qian@pnnl.gov)

Abstract. Cloud Fraction (CF) is the dominant modulator of radiative fluxes. In this study, we evaluate CF simulated in the IPCC AR4 GCMs against ARM long-term ground-based measurements, with a focus on the vertical structure, total amount of cloud and its effect on cloud shortwave transmissivity. Comparisons are performed for three climate regimes as represented by the Department of Energy Atmospheric Radiation Measurement (ARM) sites: Southern Great Plains (SGP), Manus, Papua New Guinea and North Slope of Alaska (NSA). Our intercomparisons of three independent measurements of CF or sky-cover reveal that the relative differences are usually less than 10 % (5 %) for multi-year monthly (annual) mean values, while daily differences are quite significant. The total sky imager (TSI) produces smaller total cloud fraction (TCF) compared to a radar/lidar dataset for highly cloudy days (CF > 0.8), but produces a larger TCF value than the radar/lidar for less cloudy conditions (CF < 0.3). The compensating errors in lower and higher CF days result in small biases of TCF between the vertically pointing radar/lidar dataset and the hemispheric TSI measurements as multi-year data is averaged. The unique radar/lidar CF measurements enable us to evaluate seasonal variation of cloud vertical structures in the GCMs.

Both inter-model deviation and model bias against observation are investigated in this study. Another unique aspect of this study is that we use simultaneous measurements of CF and surface radiative fluxes to diagnose potential discrepancies among the GCMs in representing other cloud optical properties than TCF. The results show that the model-observation and inter-model deviations have similar magnitudes for the TCF and the normalized cloud effect, and these deviations are larger than those in surface downward solar radiation and cloud transmissivity. This implies that other dimensions of cloud in addition to cloud amount, such as cloud optical thickness and/or cloud height, have a similar magnitude of disparity as TCF within the GCMs, and suggests that the better agreement among GCMs in solar radiative fluxes could be a result of compensating effects from errors in cloud vertical structure, overlap assumption, cloud optical depth and/or cloud fraction. The internal variability of CF simulated in ensemble runs with the same model is minimal. Similar deviation patterns between inter-model and model-measurement comparisons suggest that the climate models tend to generate larger biases against observations for those variables with larger inter-model deviation.

The GCM performance in simulating the probability distribution, transmissivity and vertical profiles of cloud are comprehensively evaluated over the three ARM sites. The GCMs perform better at SGP than at the other two sites in simulating the seasonal variation and probability distribution of TCF. However, the models remarkably underpredict the TCF at SGP and cloud transmissivity is less susceptible to the change of TCF than observed. In the tropics, most of the GCMs tend to underpredict CF and fail to capture the seasonal variation of CF at middle and low levels. The high-level CF is much larger in the GCMs than the observations and the inter-model variability of CF also reaches a maximum at high levels in the tropics, indicating discrepancies in the representation of ice cloud associated with convection in the models. While the GCMs generally capture the maximum CF in the boundary layer and vertical variability, the inter-model deviation is largest near the surface over the Arctic.

1 Introduction

Three dimensional general circulation models (GCMs) are probably the most powerful tools currently available to quantitatively investigate the Earth's climate system and to predict future climate change, which is affected by human activities that cause changes in greenhouse gases, aerosols, and land use and land cover (IPCC, 2007). From a physical point of view, anthropogenic climate change is first of all a perturbation of the Earth's radiation balance (Wild, 2008). Realistic simulation by GCMs of the perturbations of radiative forcing is an important pre-requisite for projecting reliable future climate responses. As Webb et al. (2001) emphasize: "If we are to have confidence in predictions from climate models, a necessary (although not sufficient) requirement is that they should be able to reproduce the observed present-day distribution of clouds and their associated radiative fluxes". Many previous studies have evaluated GCMs' performance in simulating shortwave (SW) and longwave (LW) radiation under cloudy and cloudless skies at the surface, where global ground radiation measurement networks are available, and/or at the top of atmosphere (TOA), where satellite observations can be used as constraints (e.g. Garratt, 1994; Wild et al., 1995, 1998, 2008; Li et al., 1997; Walsh et al., 2008). These studies found that GCMs were better at producing the mean TOA radiation budget than the surface radiation budget, although there were significant biases and inter-model variability in estimating the SW and LW radiation in particular regions (Wild and Liepert, 1998; Wild et al., 1999, 2005; Walsh et al., 2008; Wild, 2008). Although in some cases good agreement was found between the observed and modeled cloud radiative forcing, that could be a result of compensating errors in either cloud vertical structure, cloud optical depth or cloud fraction (Potter and Cess, 2004).

The climate science community has identified clouds as one of the highest priorities in climate modeling and climate change projection (IPCC, 2001, 2007). Accurate representation of cloud-radiation interactions is critical for climate models to simulate the evolution of the climate system. Clouds are also an essential variable in the climate system because they are directly associated with precipitation through microphysical processes and with aerosol loading through the aerosol aqueous-phase chemistry and wet removal process. Physically, cloud-radiation interactions depend largely on the cloud macrophysical (e.g. cloud fraction, liquid and ice water path) and microphysical (e.g. cloud droplet number, size, and ice particle habit) properties. Cloud Fraction (CF) has long been recognized as a dominant modulator of radiation flux at both the surface and the top of the atmosphere (Xi et al., 2010; Liu et al., 2011). For example, a 4 % increase in the area of the globe covered by marine stratocumulus clouds would offset the predicted 2–3 K rise in global temperature due to a doubling of atmospheric carbon dioxide (Randall et al., 1984). Although considerable uncertainties are still associated with cloud feedbacks in GCMs, one

can assume that to reasonably simulate global climate, these models should be able to accurately reproduce the current climatology of cloud fraction (including vertical structure) at a given location.

The vertical distribution of clouds affects the vertical heating rate profiles through radiative and diabatic processes, and thus influences the atmospheric stratification and general circulation (e.g. Stephens et al., 2002). Recent studies have revealed that the uncertainty in estimating the cloud occurrence at different levels is much larger than in estimating the total cloud amount in most GCMs (Stephens et al., 2002; Zhang et al., 2005; Illingworth et al., 2007; Naud et al., 2008). By comparing the results of 10 GCMs to ISCCP and CERES datasets, Zhang et al. (2005) found that models simulated a four-fold difference in high-top clouds against the observations. Because different dynamical and thermodynamic conditions produce differing vertical distributions of clouds, accurately characterizing this vertical distribution in the model is critical to understanding cloud feedback processes.

Simultaneously evaluating climatological simulations of cloud fraction, especially vertical structure, and radiation in GCMs against observations is difficult because of the lack of a long-term continuous cloud observational dataset. In situ aircraft measurements reveal the macroscopic structure of clouds, but suffer from sampling problems and can only provide 1-D cloud snapshots. Combining aircraft and ground-based instrumentation can provide a more comprehensive view of clouds and their radiative forcing, however the limitations of aircraft campaigns make this possible only for a number of isolated case studies raising the question of representativeness (Illingworth et al., 2007). Remote sensing from space has provided global cloud properties over many years (Rossow and Schiffer 1991; Webb et al., 2001), but information concerning cloud vertical structure has been lacking. The recent launch of cloud radar on CloudSat accompanied by the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO; Winker et al., 2003) provides valuable global cloud information including cloud vertical structure. However, this dataset is still relatively short, and is also limited to only two observational times per day, providing limited information on the diurnal cycle.

A review by Wild (2008) indicates that the inter-model range in SW TOA flux is about 4 % of its absolute value while the inter-model range in surface SW flux is up to 14 % of its absolute value. This result is likely due to the relative availability of global satellite versus surface data and the adjustment of model cloud parameterizations to get agreement with global mean satellite observations. Thus, information on the relationships between clouds and surface fluxes is needed to further constrain the model parameterizations so that the correct radiation budget can be obtained at both the top and bottom of the atmosphere.

In this study we use simultaneous measurements of cloud fraction and broadband radiation at the surface from the measurement sites sponsored by the Department of Energy

(DOE) Atmospheric Radiation Measurement (ARM) Program. Using radiometers, cloud radar and lidar systems, and other advanced instruments at five permanent sites located in several different climate regimes, ARM provides long-term and nearly continuous observations of the surface SW and LW radiative fluxes, sky cover and cloud vertical distributions. The long-term comprehensive ARM climatological datasets make it possible to evaluate the CF and surface radiation budgets simulated by GCMs simultaneously, which provides a unique opportunity to study the role of cloud in estimating the surface radiation budgets (Xie et al., 2010). Due to the different scales of the ARM measurements and the GCM simulations, as well as the impossibility of simulating exact weather systems in a free-running GCM, we use the multi-year ARM data to evaluate the GCMs in a climatological sense.

ARM instruments provide several CF or sky cover related products, such as cloud cover derived from the Total Sky Imager (TSI), total sky cover (TSK) derived from the surface broadband SW radiometers during daytime (Long et al., 2006), effective sky cover (ESK) using broadband LW radiometers during both daytime and nighttime (Long and Turner, 2008), and the frequency of hydrometeor occurrence statistics derived from the narrow field-of-view (FOV) lidar and radar observations (i.e. the Active Remotely-Sensed Clouds Locations or ARSCL product, Clothiaux et al., 2000). Before these data can be used to validate GCM cloud statistics, it is necessary to evaluate the measurements by these different instruments and to see if they are consistent among themselves at different time scales. One difficulty is that various observational methods and climate models use different definitions for CF (Kassianov et al., 2005). For example, the International Satellite Cloud Climatology Project (ISCCP) defines "total cloud amount" as "the fraction of the earth's surface covered by cloud" (Hahn et al., 2001). However, in surface observer climatology studies, it is defined as "the fraction of hemispherical sky covered by cloud" (Hahn et al., 2001). Climate models, on the other hand, typically interpret CF as "the horizontal area fraction covered by clouds as viewed from nadir" (Del Genio et al., 1996). Therefore, comparisons between these FOV and hemispheric observations are important if we want to utilize these data to evaluate climate models. Because of the difficulty in relating model variables to quantities retrieved from remote sensing observations, instrument simulators which use model output to directly simulate the signal that an instrument would observe have been developed in recent years. For climate models, these simulators have focused primarily on satellite observations to date. Development of techniques to simulate ground-based remote sensing observations of the type used in this study would be useful to alleviate some of the uncertainties in the model/observation comparisons.

The aim of the present paper is to evaluate the CF simulations by the GCMs in the 4th assessment report of the Intergovernmental Panel on Climate Change (IPCC AR4; IPCC, 2007), with a focus on the total cloud amount and the vertical structure of CF. Our assessment seeks to identify systematic biases and inter-model deviation in cloud fields simulated in the GCMs across seasonal scales over different regions of the world (e.g. tropics, mid-latitude continent and Arctic). Conducting the analysis over three very different climate regimes provides a better understanding of the geographical variability of clouds and their radiative forcing and a stronger constraint on model simulations.

We also examine how the CF affects cloud transmissivity, a ratio of all-sky mean downwelling SW flux to the cloud-free sky mean, in both observations and GCMs. In the following sections, we first introduce the IPCC AR4 GCMs and ARM datasets used in the analysis (Sect. 2). In Sect. 3 we compare three CF-related ARM datasets and examine their consistency at different time scales. Then we evaluate the CF simulations in GCMs against selected ARM observations, with Sect. 4 focusing on the total cloud amount and cloud transmissivity, and Sect. 5 focusing on the vertical profiles of CF. Summary and discussion are given in Sect. 6. Results of this study can help the climate modeling community to better understand the CF-related measurements from ARM sites and provide useful insights for improving the cloud-radiation interaction and the CF parameterization in climate models.

2 GCMs and ARM datasets

2.1 IPCC AR4 GCMs

The CF and radiative fluxes simulated by more than a dozen GCMs participating in the experiments for IPCC-AR4 are available from the program for climate model diagnosis and inter-comparison (PCMDI). This experimental framework is also known as the World Climate Research Program (WCRP) Coupled Model Intercomparison Project (CMIP3, Meehl et al., 2007). It should be noted that some GCMs archived both total cloud amount and CF at each model layer, but some GCMs only archived total cloud amount. There are also models that did not archive cloud-free sky surface radiation fluxes. For consistency, we use monthly mean CF and surface radiation fluxes from 11 selected GCMs in this study, as shown in Table 1. The 11 models were from NCAR (model versions ccsm3 and pcm1), GFDL (cm2), GISS (e_r), CCSR (MIROC3_2 with high resolution), MRI (cgcm2), UKMO (hadcm3), MPI (echam5), CNRM (cm3), IPSL (cm4) and INM (cm3). The model outputs used in this study are from the AMIP (Atmospheric Model Intercomparison Project) experiment, in which identical observed SSTs were used for all GCMs. Over 20 yr of results are available approximately from 1980 to 1999, with starting and ending years slightly varied among the models. More information about the project and these models can be found at the website of PCMDI (http://www-pcmdi.llnl.gov/).

Table 1. A summary of cloud fraction scheme and data availability in IPCC AR4 GCMs used in this study.

Model ID	Model shortname	Cloud Scheme	References	Total Cloud Amount	Layer Cloud Fraction	All-sky downward SW	Clear-sky downward SW
CCSM3 (USA)	ccsm	Prognostic cloud condensate with diagnostic cloud amount	Rasch and Kristjansson (1998) Zhang et al. (2003) Boville et al. (2006)	√	√	√	√
CNRM-CM3 (France)	cnrm	Statistical cloud cover scheme for stratiform clouds; Convective cloud cover based on mass-flux transport	Ricard and Royer (1993)	√		√	√
IPSL-CM4 (France)	ipsl	Diagnostic CF based on log-normal PDF of total water	Bony and Emmanuel (2001)	√	√	√	√
ECHAM5/MPI-OM (Germany)	mpi	Prognostic water phases; Bulk cloud microphysics; Relative Humidity based CF scheme	Lohmann and Roeckner (1996) Roeckner et al. (2003)	√		√	√
GFDL-CM2.0 (USA)	gfdl	Cloud microphysics after Rotstayn, 2000; Prognostic CF after Tiedtke (1993)	Rotstayn (2000) Tiedtke (1993)	√		√	√
GISS-ER (USA)	giss	Prognostic cloud condensate with diagnostic CF based on RH above a tunable RH threshold	Del Genio et al. (1996) Sundqvist et al. (1989) Schmidt et al. (2005)	√	√	√	√
INM-CM3.0 (Russia)	inm	Diagnostic CF; Stratiform CF: linear function of RH; Convective CF is 0.5 for shallow and depends on precip. for deep convection	Alekseev et al. (1998)	√	√	√	√
MIROC3.2 (hires) (Japan)	cnf_hires	Prognostic cloud water and statistical large-scale CF scheme	Le Treut and Li (1991)	√	√	√	√
MRI-CGCM2.3.2 (Japan)	mri	Diagnostic cloud water based on T Diagnostic CF based on RH, separately for convective vs layer and land vs. ocean	Yukimoto et al. (2001)	√	√	√	√
PCM (USA)	pcm	Diagnostic cloud condensate and cloud cover	Kiehl et al. (1998)	√	√	√	
UKMO-HadGEM1 (UK)	ukmo	Diagnostic with triangular PDF; Parameterized crttical RH and a vertical gradient cloud scheme	Smith (1990) Cusack et al. (1999) Smith et al. (1999)	√		√	√

CF is a critical variable in climate models for determining the radiative fluxes through the atmosphere and at the surface. Depending on the complexity of the model, CF may also be used in many other physics parameterizations in the model such as cloud microphysics, aerosol wet removal and convective transport. In this study, we focus on the role of CF in radiation, where the area-averaged CF is used. As discussed in Brooks et al. (2005), although CF produced by most cloud schemes is volume-averaged, most GCMs assume that the cloudy area of a grid box fills the entire grid box in the vertical, thus essentially assuming area-averaged CF is the same as the volume-averaged CF. In GCMs, CF can be parameterized using statistic, diagnostic or prognostic approaches. Due to space constraints, we just summarize the CF parameterization schemes for all GCMs used in this study in Table 1; for more details of each cloud scheme, including references, see http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php.

In GCMs, the vertical correlations between cloud layers have to be prescribed because cloud elements are often smaller than a typical GCM grid cell and there is no general theory for how different cloud systems should overlap (Collins, 2001). Assumptions about vertical overlap of clouds can affect the exchange of energy between the atmosphere and other components in the model, influencing not only radiative heating rates but also atmospheric temperature and hydrological processes (Collins, 2001). In the IPCC AR4 models, the most common overlap assumptions are maximum/random (Geleyn and Hollingsworth, 1979). One type of maximum/random assumption has maximum cloud overlap in each of three regions representing the lower, middle, and upper troposphere, and random overlap between these regions (e.g. Chou et al., 1998). A second type of maximum/random overlap scheme has maximum overlap between clouds in adjacent levels and random overlap between groups of clouds separated by one or more clear layers (e.g. Zdunkowski et al., 1982). The latter form of maximum/random overlap was found to be more consistent with a statistical analysis of observed cloud distributions (Tian and Curry, 1989).

2.2 ARM datasets

Our ground observational data are based on the measurements from three permanent ARM sites: the US Southern Great Plains (SGP) site in Lamont, Oklahoma, the North Slope of Alaska (NSA) Barrow site, and the Tropical Western Pacific (TWP) Manus site. These sites represent mid-latitude continent, Arctic and tropical climate regions, respectively.

ARM sites are equipped with ground-based active and passive remote-sensing instruments, including the millimeter-wavelength cloud radar (MMCR), the micropulse lidar (MPL), the laser ceilometer, broadband SW and LW radiometers, and the total sky imager (TSI). Through its value-added product (VAP) efforts, ARM has implemented advanced retrieval algorithms and sophisticated objective data analysis approaches to process and integrate data collected from these instruments (Xie et al., 2010). Below we give a brief overview of the observational datasets used in this study. More details about these ARM data products can be found in Xie et al. (2010) and in the references therein.

(a) Radiation flux and Total Sky Cover (TSK)

At the SGP site, surface radiation flux data are measured by three separate radiometer systems. An ARM Value Added Product called the Best Estimate Flux (the BEF data) (Shi and Long, 2002) combines the measurements from the three systems to produce the best estimate of surface radiation fluxes. For the NSA and TWP sites, the radiation measurements are from the SkyRad (sky radiation) and GndRad (ground radiation) systems measuring downwelling and upwelling fluxes, respectively. The total downwelling SW fluxes used in this study are primarily the sum of the direct plus diffuse components (measured by Eppley Normal Incidence Pyrheliometer and Eppley shaded Precision Spectral Pyranometers, respectively) whenever available; otherwise the global SW fluxes from the unshaded Eppley Precision Spectral Pyranometers are used. All radiation data used for this study have been quality tested using the QCRad (Quality Control for Radiation measurements) methodology of Long and Shi (2006, 2008). After the quality tests, data are further processed by the Radiative Flux Analysis (RFA). The RFA is a collection of analysis tools that detects clear-sky periods and produces continuous clear sky estimates of SW fluxes (Long and Ackerman, 2000) and LW fluxes (Long and Turner, 2008) and infers bulk cloud properties such as daylight total sky cover (TSK; Long et al., 2006), longwave effective sky cover (ESK; Durr and Philipona, 2004), and cloud effective SW transmissivity from the broadband radiometer data. These measurements have a hemispherical FOV and thus provide time series of fractional sky cover (Kassianov et al., 2005).

The TSK measurements with 1-s sampling interval have hemispheric fields of view and hence are more related to fractional sky cover, the angular amount of the sky dome covered by clouds. The derived TSK is based on measurements, so also includes the uncertainties of the measured quantities themselves. The ARM Program documents the uncertainties in broadband radiation measurements (Stoffel, 2005) as $4 \, \mathrm{W m^{-2}}$ or 3 % (whichever is greater in $\mathrm{W m^{-2}}$) for diffuse SW, $20 \, \mathrm{W m^{-2}}$ or 6 % for direct normal SW, and $4 \, \mathrm{W m^{-2}}$ or 4 % for LW. For the direct component SW, then, the uncertainty is roughly the normal incidence uncertainty weighted by the cosine of the solar zenith angle or $20 \, \mathrm{W m^{-2}}$, whichever is greater. The clear-sky estimations for diffuse, direct, and total SW are about the root mean square (RMS) of twice the measurement uncertainty for the SW components (Long and Ackerman, 2000), and about 4–$5 \, \mathrm{W m^{-2}}$ for the clear-sky LW (Long and Turner, 2008). Because the cloud effective SW transmissivity includes the instrument characteristics in both the numerator and denominator, the instrument characteristics are largely removed from the ratio. Thus the effective uncertainty of the ratio is at about the 2 % level (Long and Ackerman, 2000). For TSK, there is no "truth" for sky cover because of the nebulous definition of what is and is not a cloud in the community. However, the TSK, as well as sky imager retrievals and human sky observations all tend toward the same definition of cloud. Comparisons between TSK and sky imager and human observations give agreement to better than 10 % sky cover (Long et al., 2006). All ARM radiometer-based inferred quantities are produced at the same 1-min resolution as the measurements, and averaged to longer temporal resolution as appropriate.

(b) Total Sky Imager (TSI)

The ARM observational strategy does not include human observations, however the TSI is the instrument most similar to a traditional human observation of cloud cover. The TSI takes hemispheric "fish eye" color digital pictures of the sky every 30 s during daylight hours from a camera mounted looking down on a curved mirror. These images are then processed to infer what fraction of the sky view contains cloud elements, or fractional sky cover. The processing uses the ratio of red to blue color values for each pixel in the sky image, except for that part of the image that is masked for the camera arm and sun blocking strip on the rotating mirror. One advantage of sky imagers over human observations is consistency of the retrieved results, where the subjective nature that affects human observations is removed. An overview and examples of this processing methodology are presented in Long (2010). Comparisons with TSK give overall agreement at better than 10 % (Long at al., 2006) and with the Scripps Whole Sky Imager at the same level (Long et al., 2001).

(c) ARSCL cloud fraction

By integrating measurements from the MMCR, MPL and laser ceilometers, the ARSCL product provides an estimate

of the total amount and best estimate vertical location of clouds (Clothiaux et al., 1999, 2000). These are vertically pointing instruments with a narrow FOV that can only detect clouds directly above the instruments. However, unlike the TSI and TSK measurements, they provide both vertical location of clouds and nighttime cloud detection. The ARSCL CF is derived based on the ARSCL cloud boundary information using the algorithm described in the Climate Modeling Best Estimate (CMBE, Xie et al., 2010 and also at http://science.arm.gov/wg/cpm/scm/best_estimate.html). In this algorithm, a cloud point is first determined by MMCR or MPL and then screened by the best estimate of cloud base from both laser ceilometers and MPL to minimize the problem caused by precipitation in determining cloud base. As indicated in Clothiaux et al. (2000), the laser ceilometers and MPL can provide quite accurate cloud base measurements because they are usually insensitive to ice precipitation (if the concentration of precipitation particles is not sufficiently large) or clutter. The ARSCL CF is then calculated by averaging the cloud mask points (where cloudy, clear or missing points are set) in the one-hour time period. Therefore, the ARSCL CF actually represents the frequency of cloud occurrence rather than fractional cloud area coverage.

The cloud statistics obtained from such narrow FOV height-time transects might not be representative of a larger area surrounding these instruments at a short time scale (e.g. Berg and Stull, 2002; Kassianov et al., 2005), which is one reason we use longer-term observations for comparisons to the model results. Another issue with the ARSCL clouds is that cloud radar tends to underestimate the cloud top heights for thin high-altitude clouds because of detection limits and signal attenuation (Comstock et al., 2002). The consequence of this problem has been mitigated with the use of ARM MPL in CMBE, which is sensitive to small cloud particles. The ARSCL cloud statistics used in this study are calculated based on data during periods when both MMCR and MPL were in operation.

(d) Vertical mapping of CF

The vertical resolution (layer thickness) of the ARSCL data and the models are quite different. We vertically mapped the CF from each GCM into the ARSCL vertical grid so the model-observation comparisons presented later are all based on the finer ARSCL vertical grid. Since the CF is assumed vertically constant within each GCM grid layer, we simply distribute the CF in each GCM layer evenly into the much finer ARSCL layer (about 45 m thick). This should be similar to using a simple average to map the finer vertical resolution ARSCL CF to the coarser GCM vertical grid, but allows us to have a single vertical grid for comparison (rather than mapping ARSCL to each GCM vertical grid with different layer thicknesses).

CF and surface radiation data are available starting from late 1990s at most of the sites. Approximately 10 yr of data

are used in this study. There were frequent data gaps in the ARM observations, especially at the remote Manus and NSA sites. To make full use of the ARM datasets, we setup the following individual rules for data selections. Overall, it's required to have at least 30 % of quality-controlled data for a day or hour to be included in the analysis. For dataset inter-comparisons at daily time scale, we use the exact same periods when all involved datasets are available (e.g. for scatter and PDF Figures). For comparison of a single dataset against the GCMs at climatology scales, we include all available data from that dataset, regardless of the other two. Therefore, it is possible that different volumes of the datasets are used for different purposes.

3 Inter-comparisons of three ground-based CF related datasets

Before presenting the evaluation of the GCM results, we first present inter-comparisons of the three ground-based CF related datasets at different time scales to understand the uncertainties inherent in the observed cloud amount. To demonstrate the differences between the three datasets on a daily timescale, two time periods are selected here: 29 April to 8 July 2006 and 16 January to 26 March 2007 when all three datasets are available and both MPL and MMCR operated normally in ARSCL. Figure 1 shows the time series of daily total cloud fraction (TCF) based on ARSCL, TSI, TSK and ARSCL minus TSK over Manus, averaged only for daytime hours between 08:00 to 17:00 (local time). Overall the daytime TCF from ARSCL, TSI and TSK are temporally correlated, especially when TCF is lower (e.g. 1 and 21 June 2006) and/or wind speed is larger (e.g. 1 to 12 and 28 February to 3 March 2007). However, apparent differences can be found among them, especially between ARSCL and the other two datasets. For example, ARSCL TCF is much larger than TSK and TSI around 22–26 May 2006, as highlighted in top-left panel. This overestimation of TCF by ARSCL is probably related to persistent cloudiness located directly overhead, as shown in one of the sky images for 24 May 2006 (top-right panel) (Boers et al., 2010). An opposite case is for 17–18 March 2007, as highlighted in the bottom-left panel, in which the ARSCL underestimates the TCF probably because the clouds are visible, but not directly overhead, as shown on one of the sky images for 17 March 2007 (bottom-right panel). The wind speed is generally less than $2 \, \text{m s}^{-1}$ for the above two cases, which implies that ARSCL tends to generate larger biases in estimating the TCF as the clouds are slowly moving or stationary.

The ARM sites represent three different climate regimes, so it is important to comprehensively compare the three CF measurements over the different sites. Here we first describe the three sites and their typical meteorological conditions and then present detailed analysis at daily and monthly time scales over each site.
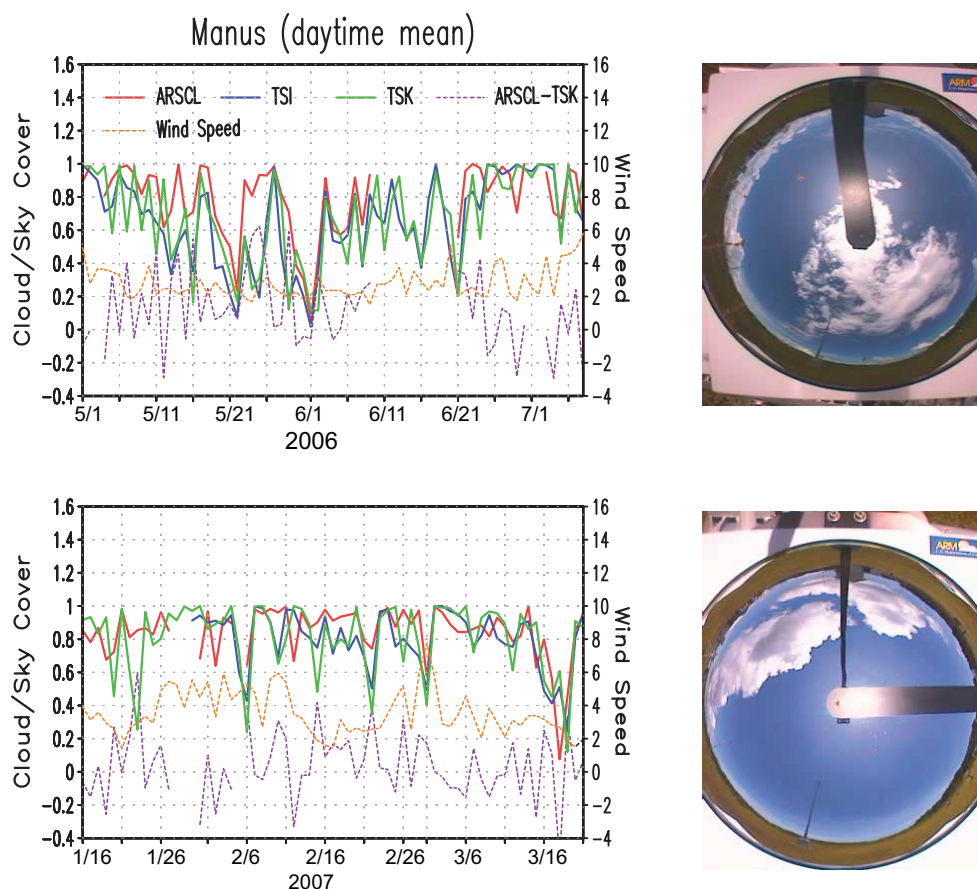
Fig. 1. The time series of daily total sky cover or cloud fraction based on ARSCL, TSI, TSK and ARSCL-TSK over Manus, averaged only for daytime hours for 29 April to 8 July 2006 (top) and 16 January to 26 March 2007 (bottom), respectively, when all three datasets are available and both MPL and MMCR operate normally in ARSCL. Right panels are examples of total sky image.

3.1 Manus

Manus, one of the three TWP sites located in the Western Pacific Warm Pool region, is influenced by the El Niño-Southern Oscillation (ENSO), which also plays a large role in the interannual variability observed in the global climate system. The TWP region consistently has warm sea surface temperatures that produce large surface heat and moisture fluxes into the local atmosphere, causing the formation of deep convective cloud systems and consequent high-altitude cirrus clouds.

The scatter plots comparing the daily averaged ARSCL, TSI, and TSK values over Manus are shown in the top panel of Fig. 2. The correlation coefficient between ARSCL and TSI is 0.63, and the root mean square deviation (RMSD) is 0.23, indicating a significant bias (>30 %) between ARSCL and TSI on the daily timescale. There is also a similar significant inconsistency between ARSCL and TSK, with a correlation coefficient of 0.56 and RMSD of 0.24 (not shown). The correlation coefficient between TSI and TSK is 0.79 and RMSD is 0.17, indicating a relatively smaller bias between TSK and TSI.

As we are interested in examining the performance of GCMs in a climatological sense, we now examine multi-year monthly means and annual means of TCF from the three datasets over each site (Fig. 3). There are approximately 10 yr of data available for ARSCL and TSK over these sites. The TSI has fewer years of observations, especially at Manus and NSA. Since only 2–3 yr of TSI data are available at NSA, we do not include it in this analysis. The average TCF ranges from 0.65 to 0.85 at Manus, with a minimum value in May. The overall seasonal variability of TCF is small over Manus. Compared to the daily-based TCF, the differences among the three multi-year averaged datasets are much smaller. The annual RMSD between ARSCL and TSI, ARSCL and TSK, TSI and TSK are all less than 0.075, indicating less than 10 % disagreement among the three multi-year averages.

Given the large differences in the daily TCF between ARSCL and TSI/TSK, it is somewhat surprising that the monthly differences are less than 10 %. Figure 4 (top) shows the frequency distribution of daily total sky cover or cloud fraction over Manus for multiple years of data. The TSI frequency is larger (smaller) than ARSCL when CF is less (larger) than 0.6. Also the difference between the TSI and
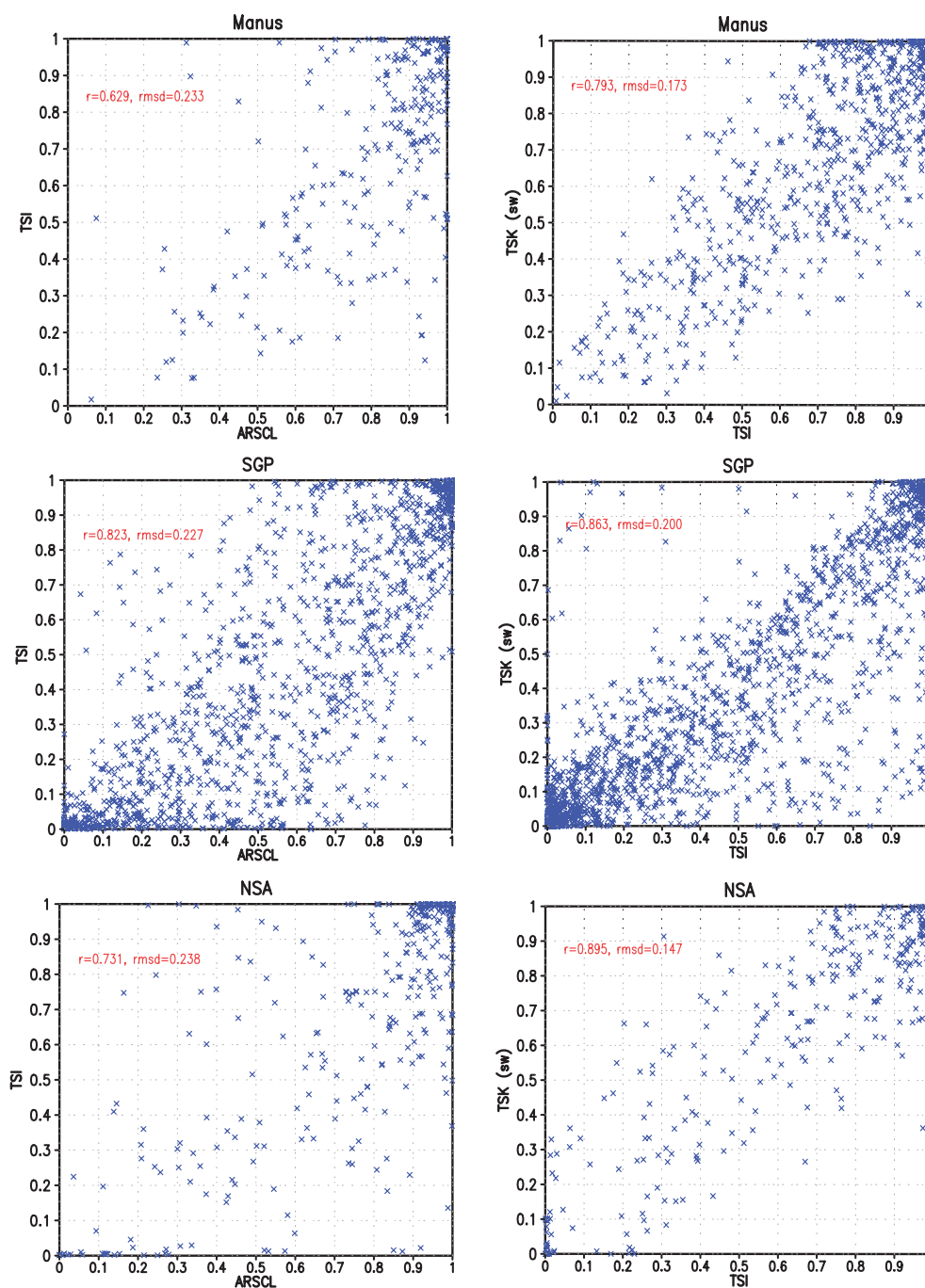
Fig. 2. The scatter plots for daily total sky cover or cloud fraction based on ARSCL and TSI (left) and TSI and TSK (right) over Manus (top), SGP (middle) and NSA (bottom). Only the days when both datasets are available are included in each plot. At NSA only days from April to September (when solar elevation is high enough for reliable TSI and TSK measurements) are used. For each panel, the correlation coefficient and root mean square deviation (RMSD) are given.

ARSCL tends to be significant when CF is larger than 0.8 or smaller than 0.2. This is not surprising due to the several orders of magnitude difference in field-of-view between the TSI and ARSCL. During days with large amounts of CF, because of the small sampling of the ARSCL narrow FOV, the

ARSCL beam is more likely to be filled with cloud and hence overestimates the TCF. On days with lower cloud cover, the ARSCL beam is more likely to sample clear sky than cloud, and hence underestimates the TCF. The overestimation for larger CF by ARSCL is compensated by less frequent smaller
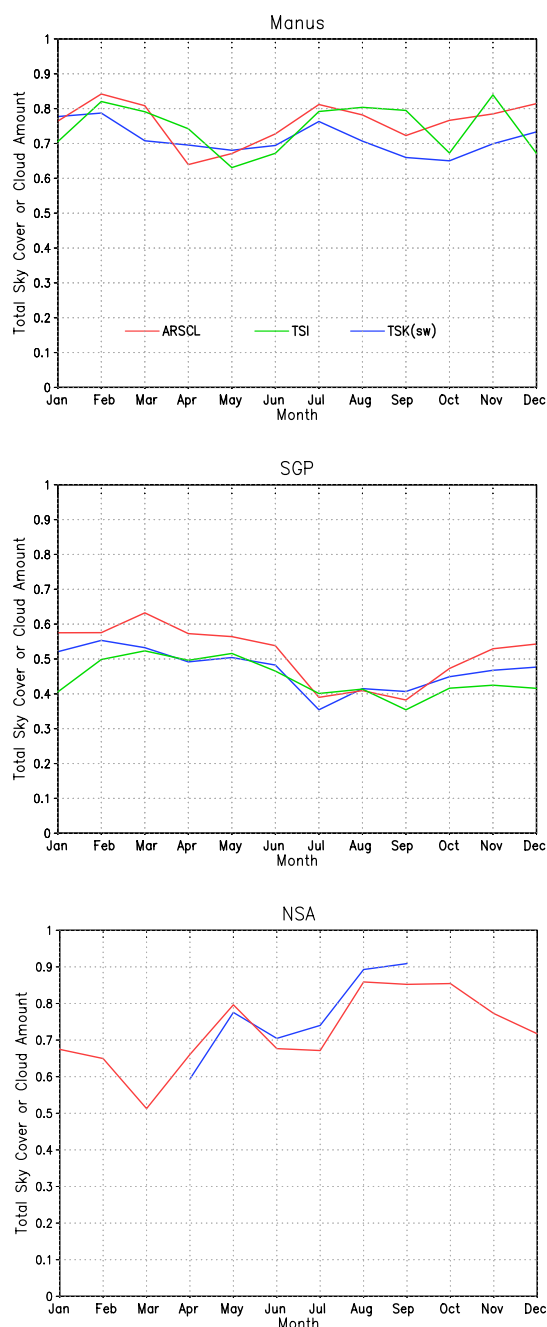
Fig. 3. The multiple-year averaged monthly mean total sky cover or cloud fraction from ARSCL, TSI and TSK over Manus (top), SGP (middle) and NSA (bottom), respectively.

CF, resulting in the small difference among the multi-year monthly mean CFs. The difference in the frequency distributions between TSI and TSK are smaller because both of them have hemispheric fields of view.

3.2 SGP

The SGP site is located in north-central Oklahoma, representing the interior regions of many mid-latitude continents, where the clouds are driven by frontal systems or by heating and local convection. The convection is usually short lived over the SGP and does not have the extensive cirrus that is found in the tropics (Mace and Benson, 2008). Shallow cumuli often form in spring and summer under stable synoptic conditions with a strong surface forcing and well-developed boundary layers (Berg et al., 2010).

The scatter plots for the daily ARSCL and TSI, and TSI and TSK values over SGP are shown in the middle panel of Fig. 2. The correlation coefficient between ARSCL and TSI is 0.82, and the RMSD is 0.23, indicating a significant bias between ARSCL and TSI on the daily timescale. The correlation coefficient between TSI and TSK is 0.86 and RMSD is 0.20. The differences among three measurements as shown from scatter plots are similar at SGP with that at Manus.

Figure 4 (middle) shows the frequency distribution of daily total sky cover or cloud fraction over SGP. While the occurrence frequency of cloudiness shows an upward trend with the increase of CF at Manus, much more clear sky days can be found at SGP than at either NSA or Manus, where most of the time there are clouds at least part of the day. The TSI frequency is larger (smaller) than ARSCL when CF is less (larger) than 0.3. The TSI is more than two times larger than the ARSCL when CF is less than 0.1. Similar to Manus, the difference of frequency distribution between TSI and TSK are much smaller than between TSI and ARSCL.

The multi-year mean TCF over SGP ranges from 0.35 to 0.62 (Fig. 3, middle). The overall magnitude of TCF is significantly smaller than over the tropics, although the seasonal variability is greater. The maximum TCF is during winter and spring and the minimum is during July to September. Similar to the results seen at Manus, the monthly mean TCF from ARSCL is larger than that from TSI or TSK, but disagreement among the three datasets is less than 15 %.

3.3 NSA

The NSA site is located at Barrow, the northernmost location in Alaska. This site, located near cryospheric boundaries, has a prevailing east-northeast wind off the Beaufort Sea and is influenced by both extratropical and Arctic synoptic activity (Stone et al., 2002). Previous research has estimated that clouds in the Arctic are more prevalent and persistent than clouds elsewhere (Curry et al., 1996). Additionally, the Arctic site has a large amount of mixed-phase clouds, which are not well treated in current climate models (Verlinde et al., 2007).

The bottom panel of Fig. 2 shows the scatter plots for the daily ARSCL and TSI, and TSI and TSK values over NSA. At NSA only days from April to September (when solar elevation is high enough for reliable TSI and TSK

measurements) are used in this analysis. The correlation co-efficient between ARSCL and TSI is 0.73, and the RMSD is 0.24, similar to that at the other two sites. The correlation coefficient between TSI and TSK is 0.90 and RMSD is 0.15, indicating a better agreement between TSI and TSK than at the other two sites.

Figure 4 (bottom) shows the frequency distribution of daily total sky cover or cloud fraction over NSA. Similar with at Manus, the occurrence frequency of cloudiness shows an upward trend with the increase of CF and overcast skies (CF > 0.9) account for almost 40 % of days at NSA. Generally, the TSI frequency is larger (smaller) than ARSCL when the CF is less (larger) than 0.8. Similar to the other two sites, the difference of frequency distribution between TSI and TSK are smaller than between TSI and ARSCL.

The multi-year mean ARSCL TCF ranges from 0.5 to 0.9 at NSA, showing a stronger seasonal variability over the Arctic than at Manus or SGP. TCF increases significantly from March to May (0.5 → 0.8), remains relatively high from May to October except for June and July, and then decreases from October to the next March. The maximum TCF occurs in August–October and the minimum occurs in March. For the available months, the ARSCL and TSK match very well and the difference between them is less than 10 %.

3.4 A summary and discussion for inter-comparison of three datasets

For all sites, the correlation coefficients are higher and RMSD are lower for the TSI/TSK comparisons than for the ARSCL/TSI comparisons. This is not surprising because ARSCL is derived from the time-slice measurement with narrow lidar/radar FOV, but both TSI and TSK are from hemispheric observations. There are several possible reasons for the bias between the daily TCF from ARSCL and the TSI/TSK measurements. The first possibility is the different fields of view. Although one might expect the TSI/TSK to have a higher CF than ARSCL because the hemispheric FOV instruments are more likely to be affected by cloud sides as well as cloud bases, it is also true that the narrow FOV ARSCL instrument samples only a very small fraction of the domain seen by the hemispheric view instruments (TSI and TSK). Thus, if the cloud field is not isotropic, then sampling such a small portion of the cloud field could easily lead to large biases in CF over short time periods. Pincus et al. (2005) used cloud scenes produced by a 3-D large-eddy simulation model to simulate the CF that would be seen by a vertically pointing narrow FOV instrument and compare it to the model's domain-mean CF. They found that the difference in cloud fraction varied from scene to scene and also depended on the averaging period used.

Another reason for differences in the CF is that the TSI, broadband radiometer, and radar/lidar measurements use very different techniques to detect cloud and thus have different sensitivities to different types of clouds. The MPL

Table 2. Annual mean Total Cloud Fraction (TCF) over three sites.

Site	ARSCL	TSI	TSK	ARSCL or other instruments				
Manus	0.76	0.74	0.71					
SGP	0.51	0.45	0.47	0.47[1]	0.49[2]	0.49[3]	0.50[4]	0.50[5]
NSA	0.73		0.77[6]	0.78[7]	0.75[8]			

[1] ARM SGP Radar-lidar, 1997–2006, Table 2 in Xi et al. (2010).

[2] ARM SGP Radar-lidar, 1997–2002, Dong et al. (2006).

[3] ARM SGP ARSCL, 1998–2004, Kollias et al. (2007).

[4] Global 5×5° surface observation from synoptic weather stations for 1971–1996, Warren et al. (2007).

[5] Surface observation from two synoptic weather stations nearby SGP for 1981–1991, Lazarus et al. (2000).

[6] Only averaged from April to September.

[7] ARM NSA Radar-lidar, Table 2 in Dong et al. (2010).

[8] ARM NSA Ceilometer, Table 2 in Dong et al. (2010).

instrument, which is included in the ARSCL TCF, can detect very optically thin cirrus clouds that may not significantly affect the broadband SW measurements used to determine the TSK cloud amount. The significant bias between narrow FOV and hemispheric observations on a daily basis suggests that users should be extremely cautious to use these datasets to quantitatively evaluate the hourly or daily CF calculated in climate models or retrieved by satellite instruments.

The Manus and NSA sites both have large frequency of overcast cases and relatively few clear sky cases compared to SGP. At Manus, much of the overcast is likely due to ice anvil and cirrus associated with deep convective systems, while at NSA there is often extensive low-level cloudiness. At all sites, the ARSCL frequency is less than TSI when CF is small (<0.3) and greater than TSI when CF is large (>0.8). Also the difference between the TSI and ARSCL tends to be significant when CF is larger than 0.8 or smaller than 0.2. The compensating errors in lower and higher CF days result in small biases of TCF between ARSCL and TSI/TSK measurements as multiple-year data is averaged.

Table 2 summarizes the annual mean TCF from the different observations and some previous studies over the three sites. The mean TCF over Manus derived from the ARSCL, TSI and TSK is 0.76, 0.74, and 0.71, respectively. We did not find any previous studies that summarized annual mean TCF over Manus. The TCF ranges from 0.45 to 0.51 over SGP, based on this and some previous studies. The TCF from ARSCL in SGP is 0.51 in this study, very close to the values (0.49–0.50) from other studies, including those from synoptic weather stations. The annual TCF based on ARSCL are larger than 0.73 over NSA, which are comparable to those derived from ground-based radar/lidar observations during the Surface Heat Budget of the Arctic Ocean experiment and from satellite observations over the western Arctic regions (Walsh et al., 2009).

Although the point-to-point differences are significant among the three CF datasets on a daily basis, the differences
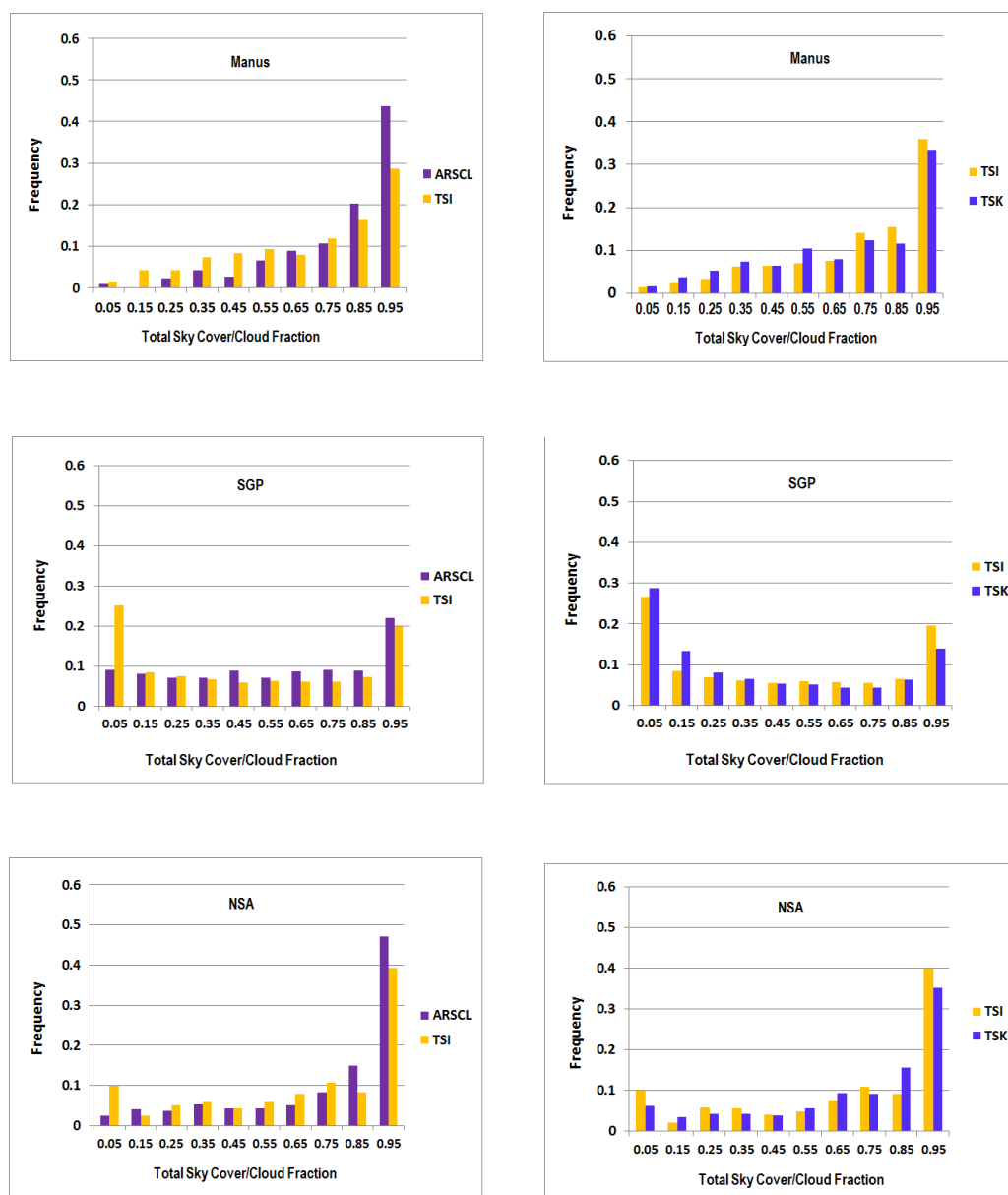
Fig. 4. The frequency distribution for daily total sky cover or cloud fraction (left: ARSCL vs. TSI; right: TSI vs. TSK) over Manus (top), SGP (middle) and NSA (bottom).

are less than 15 % in the multi-year monthly means and less than 5 % in the annual means over the three sites. The differences in annual mean TCF among ARSCL, TSI, TSK and observation from synoptic weather stations are less than 0.04 (8–9 %) over SGP. Additionally, it gives us some confidence in using the narrow field of view ARSCL data to assess the models' vertical distribution of clouds. Thus the multi-year monthly means are the best available dataset to compare to the climate model results. However, this does not necessarily mean that the estimates of measured CF are unbiased, only that averaging over a longer time period and a multi-

plicity of cloud types tends to offset detection differences between the different instruments.

Meanwhile, Xi et al. (2010) analyzed one decade of ARM ARSCL and Geostationary Operational Environmental Satellite (GOES) observations at the SGP site and revealed an excellent agreement in the long-term mean CF derived from the surface and GOES data. Dong et al. (2006), Xi et al. (2010), and Kennedy et al. (2010) have also found ARSCL CF to be statistically representative in long-term monthly or annual averages of the entire sky when compared with long-term satellite and surface observations, suggesting that the

long-term ARM point observations can represent large areal observations. The consistency between long-term mean narrow FOV, hemispheric and satellite observations provides confidence for using ARM multi-year averaged monthly data to evaluate CF in climate models.

4 Evaluation of surface radiation flux and total cloud amount

4.1 Inter-model divergence and model biases

Here the TCF and the surface SW fluxes simulated in the major GCMs in IPCC-AR4 are inter-compared, and they are also evaluated against ARM measurements over all three sites. We compare the SW radiative fluxes under both cloud-free and cloudy skies and also calculate the cloud radiative effects and cloud transmissivity, attempting to investigate the role of TCF and other dimensions of cloud in contributing to the biases of simulated solar radiation fluxes in the GCMs. We find that the performance of GCMs in simulating radiative fluxes is highly related to their simulation of CF. Positive biases in monthly surface downwelling SW flux can be found when the CF is underestimated (figure not shown). However, our focus here is to compare the 11 GCMs as a group in estimating CF and related radiative effects rather than to examine the performance of each individual model.

Figure 5 compares the aggregate normalized standard deviation (NSD) of annual mean surface downward SW radiation under clear skies (CSWdn) and all-skies (SWdn), cloud transmissivity (SWdn/CSWdn, TRANS), TCF (TSK, because it matches the surface radiation flux data), and cloud effect (CSWdn-SWdn normalized by TCF) over three sites for the 11 GCMs and for the difference between models and measurements, respectively. We define NSD as

$$\mathrm{NSD} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}}{\bar{x}},$$

where x_i represents annual mean value of a variable in a GCM, and N is the number of GCMs used for calculation. $\bar{x}$ represents an observational value or the average of x_i for the 11 models,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Figure 5 (top) shows that the inter-model NSD for CSWdn is less than 0.03 (3 %), and the model-measurement NSD for CSWdn is also small (less than 0.04), indicating that the models generally do a reasonably good job with surface SW radiation in cloudless skies. The sites examined generally have small aerosol optical depths which may partly contribute to the good agreement between model and measurement for the cloudless sky fluxes.
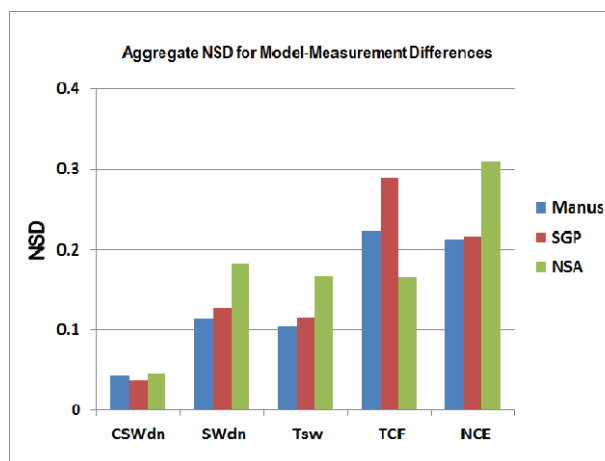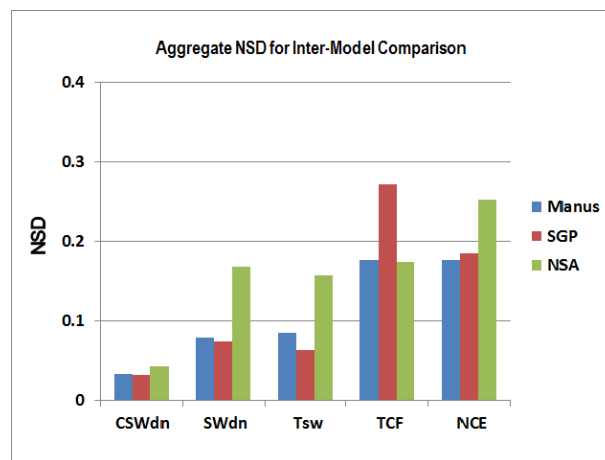

Fig. 5. The aggregate normalized standard deviation (NSD) of annual mean surface downward solar radiation under cloud-free skies (CSWdn) and all-skies (SWdn), cloud transmissivity (Tsw, SWdn/CSWdn), total cloud fraction (TCF), and cloud effect (NCE, CSWdn-SWdn normalized by TCF) over Manus, SGP and NSA, for inter-model comparison within 11 GCMs (top) and model-measurement difference (bottom), respectively.

The inter-model NSD for all-sky SWdn is around 0.08–0.18, indicating around 3–5 times larger disparity among the GCMs in estimating the downward solar radiation when clouds are included. The NSD for cloud transmissivity is very close to the value for SWdn over all three sites, because the deviation in transmissivity is mainly contributed by the SWdn rather than CSWdn. At Manus and SGP, the inter-model NSD for TCF reaches 0.18–0.28, 2–3 times as large as for SWdn and transmissivity, which indicates the inter-model disparity in TCF is much larger than in downward solar radiation. Meanwhile, the NSD of normalized cloud effect (NCE), defined as (CSWdn-SWdn)/TCF, shows a similar magnitude of NSD as TCF, indicating that other dimensions of cloud in addition to cloud amount, such as cloud optical thickness

and/or cloud height, have a similar magnitude of disparity as TCF within the GCMs. This also suggests that the better agreement among GCMs in the cloudy-sky SW fluxes than in TCF or NCE could be a result of compensating effects from errors in cloud vertical structure, overlap assumption, cloud optical depth and/or cloud fraction.

The NSD for model-measurement comparison shows a similar overall pattern for inter-model deviation, but the magnitude is slightly larger for all quantities (Fig. 5, bottom). The NSD for SWdn and transmissivity is 3–4 times as large as for CSWdn, and the NSD for TCF and NCE is twice as large as for SWdn and transmissivity, except for NSA where only six months of data from April and September are used. The similar overall deviation pattern between inter-model and model-measurement comparisons suggests that the climate models tend to generate larger differences against observations for those variables with larger inter-model deviation. The model-measurement NSD values for CSWdn, SWdn, and SWdn/CSWdn are similar at both Manus and SGP. At NSA, however, both inter-model deviation and model-measurement difference have similar magnitudes for TCF, SWdn and Tsw, suggesting that models have more difficulties in simulating surface radiative fluxes in high-latitude regions.

4.2 Comparison of seasonal cycle of total cloud fraction

Figure 6 shows comparisons of seasonal TCF for the 11 GCMs and their averages with the three different observations over Manus, SGP and NSA. Although overall seasonal variability of TCF is small over Manus, most of the models capture the minimum of TCF during the late spring. Simulated TCFs are more scattered during June to October than other months. While the simulated TCF are very diverse among the GCMs, the ensemble mean TCF (averaged over all the 11 GCMs) is close to the measurement in both magnitude and seasonal cycle.

For the long-term average, most of the GCMs (except for one model) capture the seasonal variation of TCF over the SGP, with a maximum during winter and spring and a minimum during July to September. Compared to the measurements, most of the GCMs underestimate the TCF, so the 11-model average of TCF is consistently smaller than the observations by 0.05–0.1 for almost all months.

As shown in Fig. 6, the NSA Barrow site has a relatively large cloud fraction, especially during the warm season in which the low-level cloud is persistent. This persistent large cloud coverage insures its important role in the Arctic climate system. Unfortunately, the performance of GCMs is very diverse over NSA, especially for cold season. The 11-model averaged TCF is close to the observations, with the exception of months from January to April, during which most of the models overpredict the CF by up to 0.4.
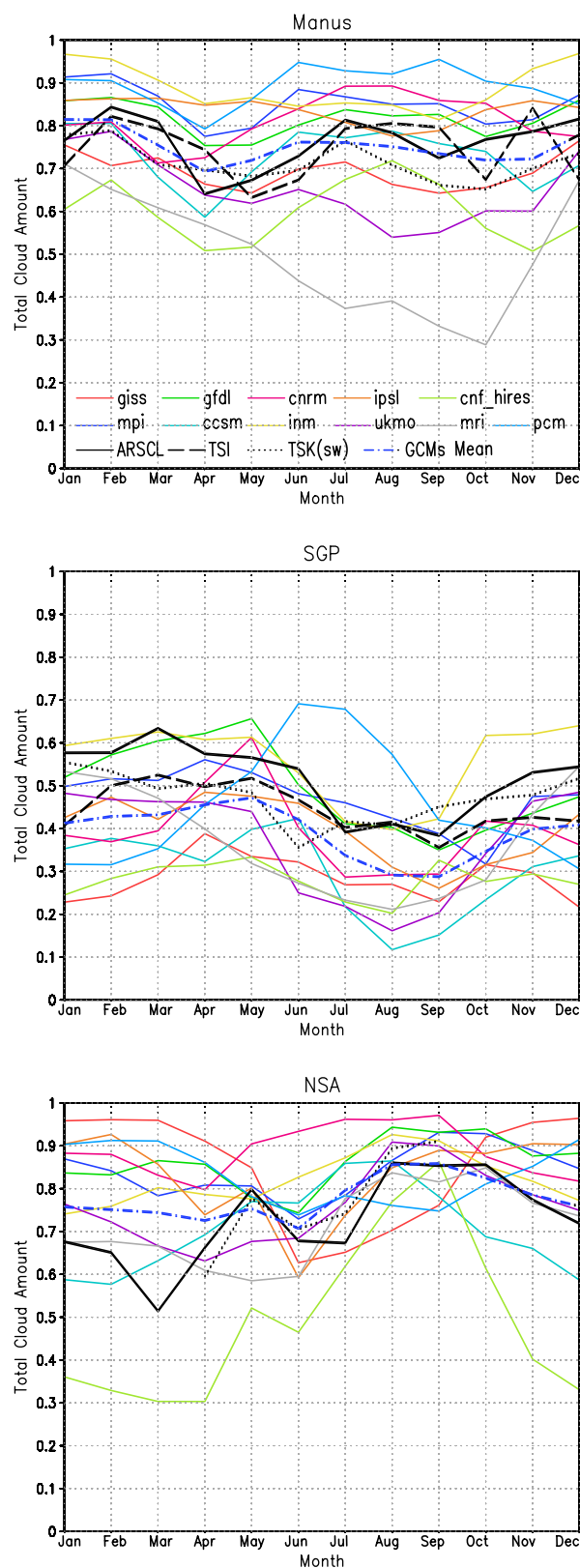

Fig. 6. The multi-year averaged monthly mean TCF for 11 individual GCMs and their average along with three different observations over Manus (top), SGP (middle) and NSA (bottom).
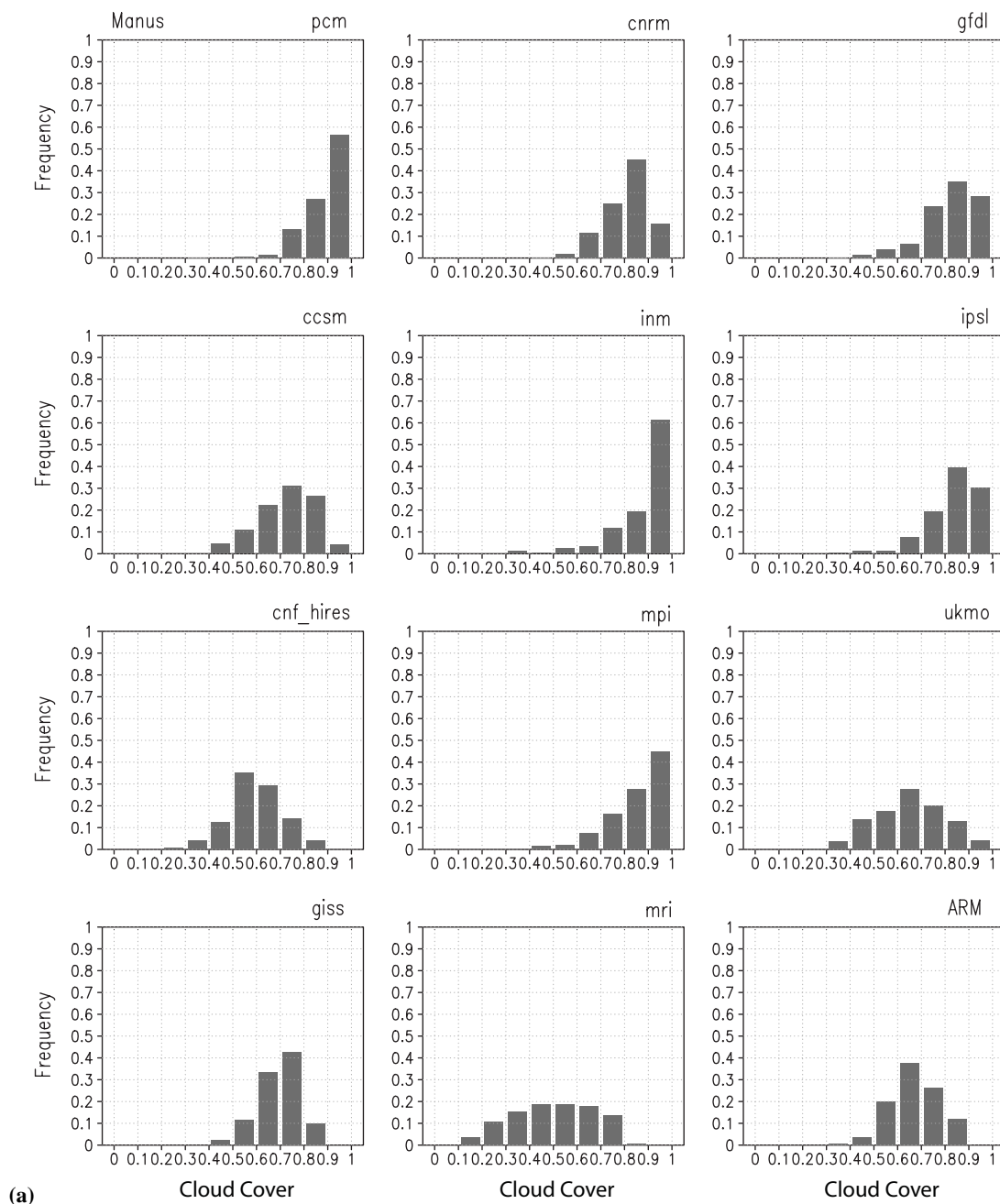
Fig. 7. The frequency of occurrence (PDF) of monthly mean TCF simulated by 11 GCMs and calculated from observation over three ARM sites. **(a)** Manus, **(b)** SGP, and **(c)** NSA.

4.3 Frequency of occurrence vs. TCF bins

Figure 7 shows the frequency of occurrence of monthly mean TCF simulated by the 11 GCMs and calculated from observations (TSK, which has longest record to calculate PDF) over the three ARM sites. Over Manus (Fig. 7a), the observed TCF shows a narrow nearly normal distribution, with a range from 0.4 to 0.9. Four GCMs, i.e. GISS, CCSM UKMO and CNF-Hires, generally well capture the observed range and

PDF pattern of TCF. The PDF pattern from CNRM, GFDL, IPSL is slightly skewed to high TCF compared to the observed nearly normal distribution, indicating a too frequent larger cloud cover in those models. Simulated frequency of occurrence in PCM, INM and MPI dramatically increases from lower to higher TCF bins with approximately 60 % of the occurrences having TCF larger than 0.9 in PCM and INM and 45 % of the occurrences having TCF larger than 0.9 in MPI. Apparently too frequent nearly overcast days are

(b)
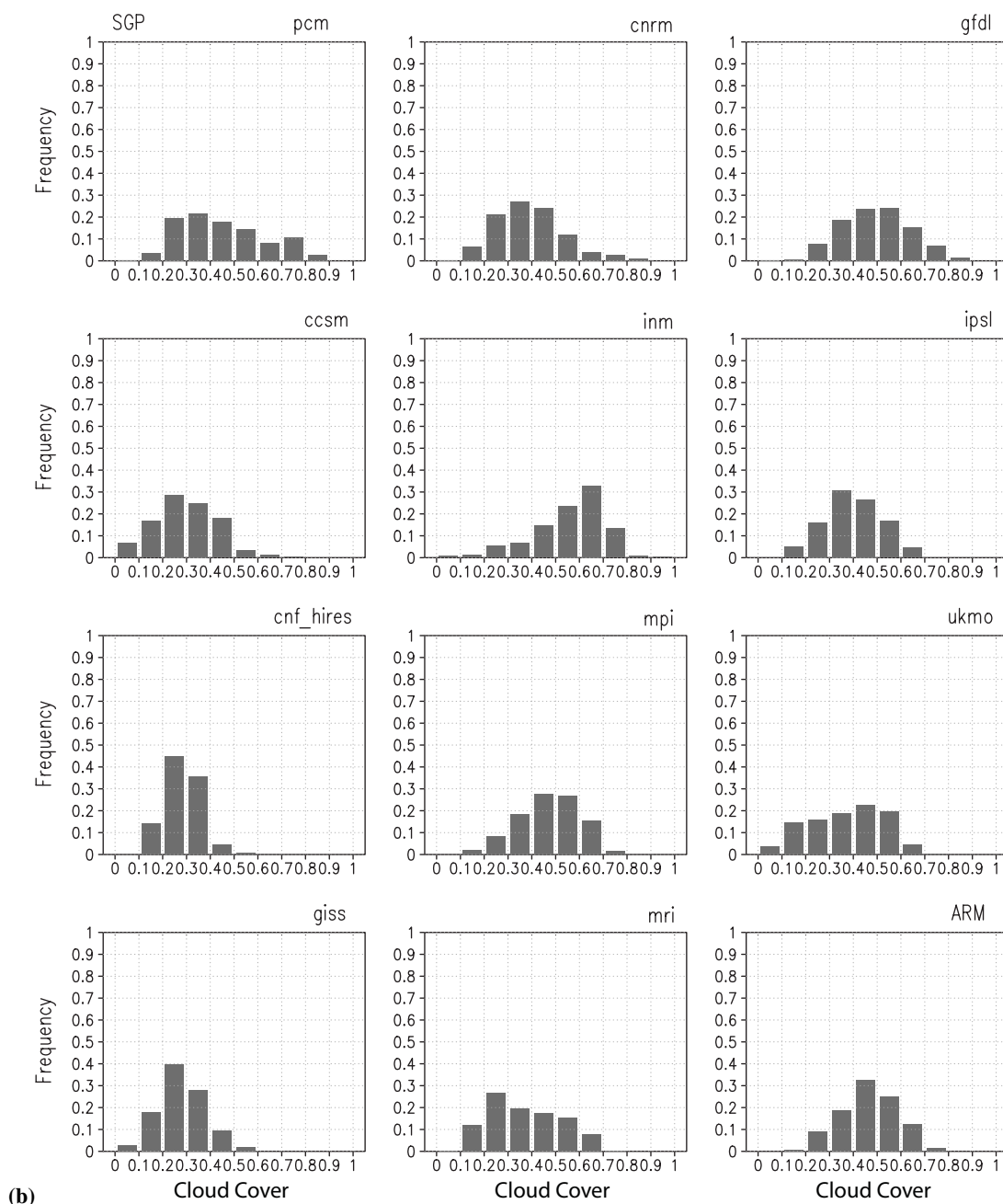
Fig. 7. Continued.

simulated over the tropics in these three models. TCF in MRI is almost evenly distributed in bins between 0.1 to 0.8, indicating a too frequent overprediction of low CF and underprediction of larger CF in this model.

Over SGP (Fig. 7b), the observed TCF also shows a near-normal distribution, with the mean about 0.5. The PDF of TCF at SGP has a similar shape to the PDF at Manus, but values are shifted to lower bins. A few of the GCMs reasonably capture the near-normal distribution pattern of TCF over SGP, such as IPSL, GFDL, MPI and CNRM. However,

the simulated TCF in PCM and UKMO are too evenly distributed over a bigger range of values. In contrast, TCF is too narrowly distributed in CNF_Hires. This model together with CCSM and GISS all show a shifted TCF distribution to the lower bins, indicating too many cloud-free and/or low cloud-cover days in these three models. Overall the GCMs perform better at SGP than at Manus in terms of the frequency distribution. This is likely related to the weaker large scale forcing at the TWP compared to SGP, and/or to the more diverse and varying cloud regimes over the tropical region, such as
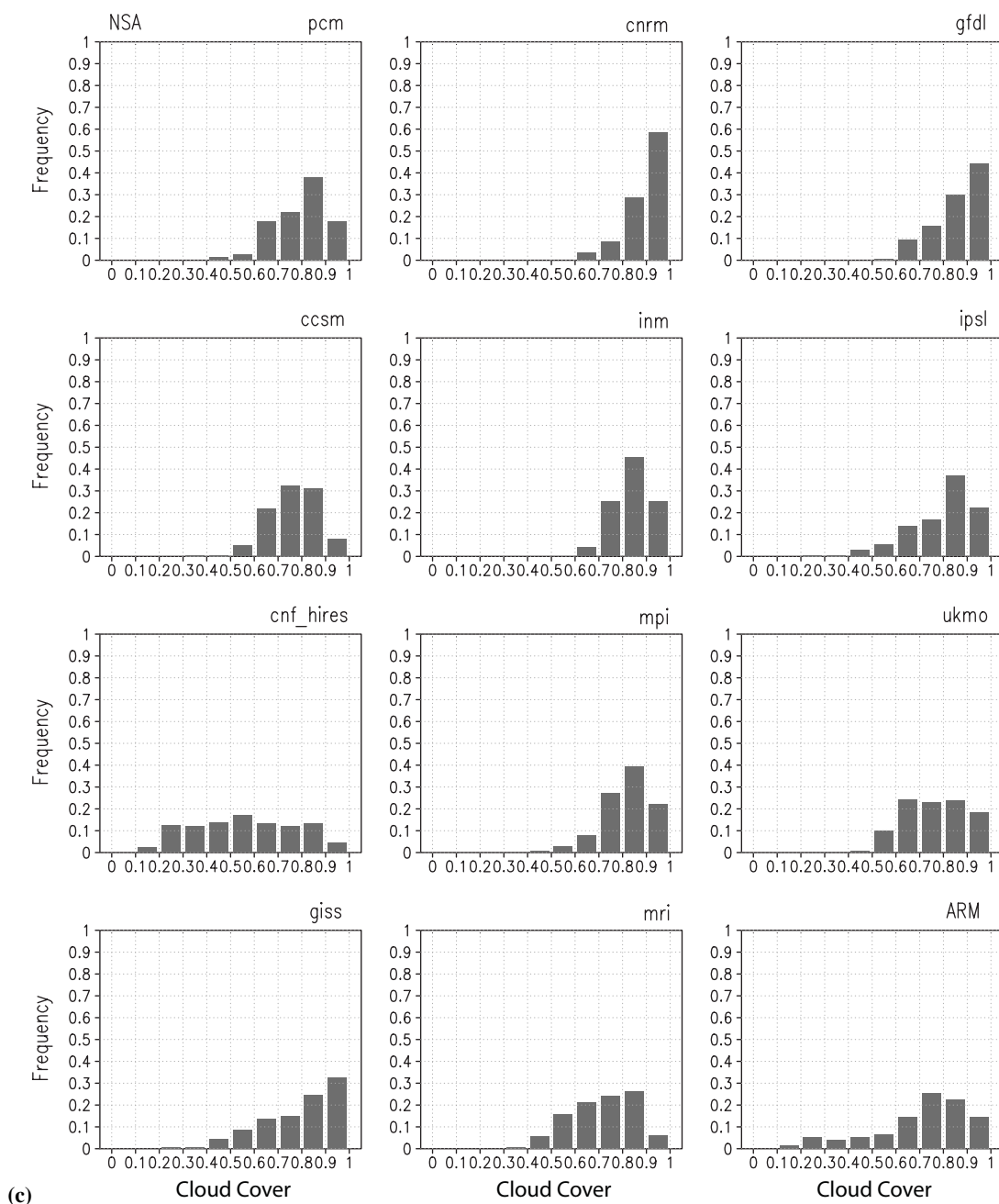
Fig. 7. Continued.

frequent deep convection and cirrus cloud, which are relatively poorly simulated in GCMs.

Over NSA (Fig. 7c), observed TCFs are widely spread in bins between 0.1 and 1.0 but occur more frequently in the higher bins. Except for CNF_Hires, all models tend to show a more narrowly distributed PDF that skews to the higher side of the bins, especially CNRM, INM and MPI, which means too many overcast and/or high cloud-cover days but too few low cloud-cover days in those models.

4.4 Transmissivity vs. TCF

Not only the total cloud cover but also cloud optical properties influence the amount of SW flux reaching the surface. The absolute impact of the cloud also depends on the magnitude of incoming solar radiation. To characterize the normalized impact of clouds on the surface SW radiation, we use the SW cloud transmissivity (Tsw, or SWdn/CSWdn). It is normalized by the clear-sky downwelling flux at the surface instead of at the top of the atmosphere so that the

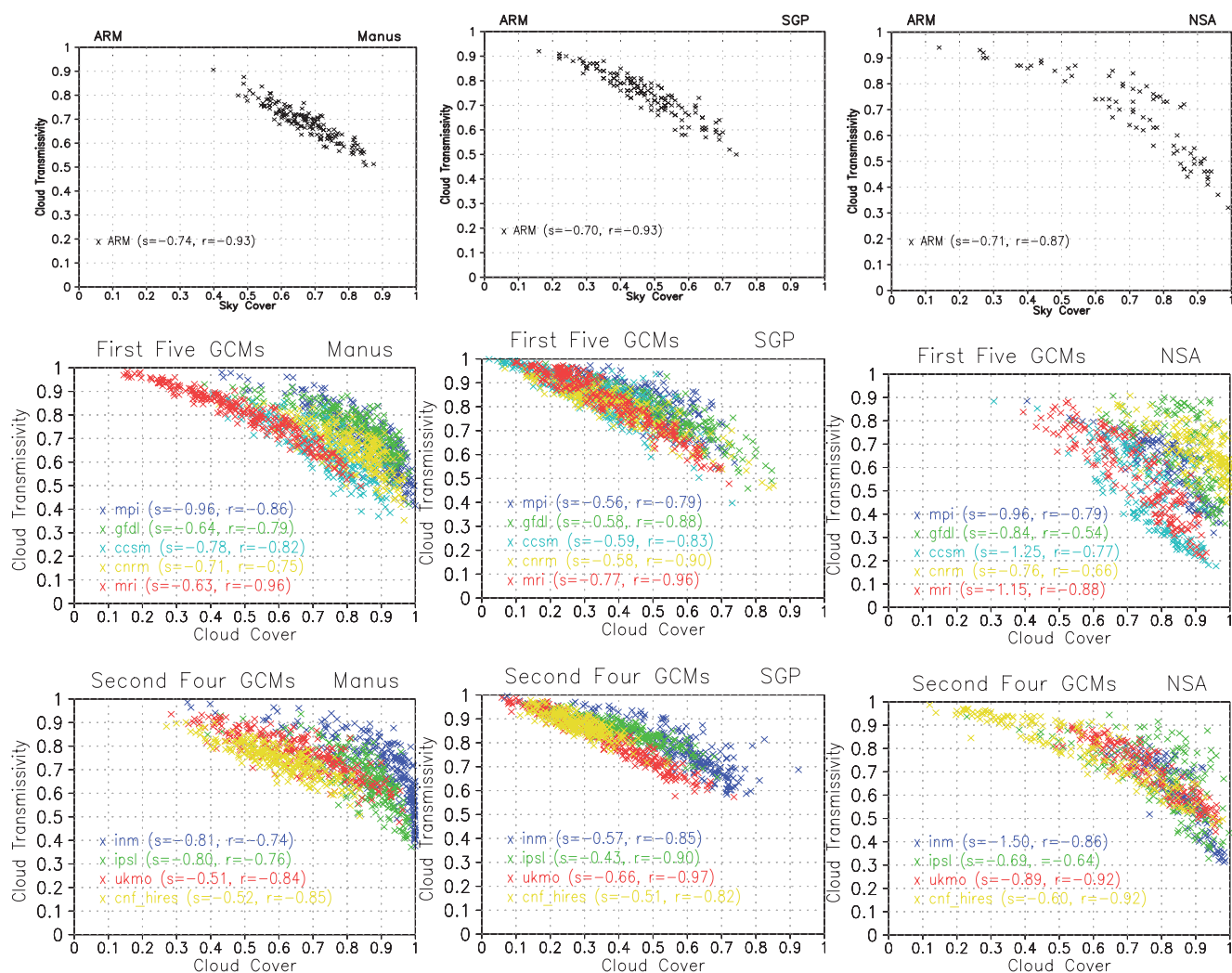**Fig. 8.** The monthly mean SW transmissivity (Tsw) against their corresponding TCF for both observation and nine GCMs over three ARM sites. **(a)** Manus (left), **(b)** SGP (middle), and **(c)** NSA (right).

effect of the atmosphere on the surface fluxes is minimized and then model treatment of molecular scattering, gaseous absorption, and aerosol is less important to the results (except for potential aerosol indirect effects). The monthly mean SW transmissivity is plotted against the corresponding TCF in Fig. 8 for both the observations and nine GCMs over the ARM sites (the other two GCMs did not archive SW fluxes under cloud-free skies). Since the SW flux under cloud-free skies (i.e. CSWdn) is much better simulated and has less scatter among the GCMs (see Fig. 5), the performance of models in estimating the transmissivity primarily reflects their ability to estimate the cloud influence on the SW flux under all-skies (i.e. SWdn). Figure 8a for Manus shows that the observed transmissivity ranges from 0.5 to 0.9. The observed transmissivity and TCF are highly correlated over Manus, with a correlation coefficient of -0.93. Moreover, the transmissivity almost linearly decreases with

increased TCF within the range of observed TCF. The slope of the fitted line (i.e. $s = \Delta T_{\mathrm{SW}}/\Delta \mathrm{TCF}$) is -0.74, and serves as an indicator of how the aggregate cloud optical properties change with changing cloud amount. For the models, the linear fit slope serves to indicate, given the cloud amounts that the model produces, whether the resultant aggregate cloud optical properties are in line with the observations for that climate regime.

The correlation coefficients between transmissivity and TCF over Manus simulated by the nine GCMs range from -0.74 to -0.96, with most GCMs having less correlation between transmissivity and TCF than seen in the observations. The simulated transmissivity, ranging from 0.4 to 1.0, is overall nominally consistent with the observed. However, the slope (or $\Delta T_{\mathrm{SW}}/\Delta \mathrm{TCF}$) varies from -40.51 to -0.96, indicating a very wide range of different cloud transmissivity changes per unit TCF change among the GCMs.

For example, the $\Delta T_{SW}/\Delta TCF$ for UKMO and CNF_Hires are -0.51 and -0.52, respectively, and the smallest values among all GCMs. Both these models exhibit an underestimation of TCF (Fig. 6a) with larger transmissivities at the lower TCF range. At the same time, these models overestimate the transmissivity for larger TCF compared to the observations. These differences suggest that the cloud optical thickness is underestimated in these models, resulting in a smaller $\Delta T_{SW}/\Delta TCF$ than observed. The $\Delta T_{SW}/\Delta TCF$ is also underestimated in MRI, which also has too many lower TCF values, and in GFDL, which has too many large TCF occurrences. However, $\Delta T_{SW}/\Delta TCF$ is -0.96 in MPI, which is much higher than the observations or other GCMs. Here again, the MPI TCF frequency is biased toward large TCF, with anomalously high transmissivity for the few occurrences of TCF in the 40–60 % TCF range. This indicates that the transmissivity in MPI is too optically thin for the mid-TCF values, however this is compensated for by the overestimated TCF in MPI (see Fig. 6a) and still results in a reasonable estimation of surface SW flux. Other models show a more reasonable agreement for $\Delta T_{SW}/\Delta TCF$ over Manus. The slope of transmissivity against TCF and the correlation coefficients between them for the observations and all GCMs and are summarized in Table 3.

Over SGP, the observed transmissivity ranges from about 0.5 to 0.9, similar to that over Manus and the correlation coefficient between TCF and transmissivity is the same as at Manus. The $\Delta T_{SW}/\Delta TCF$ is -0.70, slightly smaller than at Manus. Except for MRI, all GCMs generally underestimate the $\Delta T_{SW}/\Delta TCF$ over SGP (however MRI had a large underestimate of TSW/TCF at Manus). The minimum $\Delta T_{SW}/\Delta TCF$ is -0.43 in IPSL, almost 40 % smaller than the observed. As discussed in Sect. 4.2, most of the GCMs significantly underestimate the TCF over SGP. This indicates that current global models tend to remarkably underpredict both TCF and $\Delta T_{SW}/\Delta TCF$ over SGP, i.e. cloud cover is smaller and the models tend to generate larger transmissivities at the larger TCF values. For instance, the observations give transmissivity values ranging from 0.5 to 0.7 for TCF from 0.6 to 0.7. Yet the models (except for MRI) range from about 0.6 to 0.7 for the same range of TCF, which then produces the underestimation of $\Delta T_{SW}/\Delta TCF$.

Over NSA, a greater variety of $\Delta T_{SW}/\Delta TCF$ can be found for both observations and models (Fig. 8c). This is at least in part due to the bi-modal behavior of the relationship for snow covered and non-snow-covered ground. In the snow covered case, multiple reflection of SW between the surface and the clouds increases the SWdn, which increases the SWdn/CSWdn ratio and the ratio then includes not only the actual cloud transmission but also the multiple reflection. The snow covered ground cases are those in the upper right of the observations, with both the snow-covered and non-snow-covered cases producing about the same $\Delta T_{SW}/\Delta TCF$ slope. The transmissivity varies from 0.2 to 0.9 in GCMs for TCF larger than 0.9, indicating more divergence in transmissiv-

Table 3. Slope and correlation coefficient (in parentheses) between cloud cover and cloud transmissivity based on ARM data and nine GCMs results over three ARM sites.

	Manus	SGP	NSA
ARM	$-0.74(-0.93)$	$-0.70(-0.93)$	$-0.71(-0.87)$
cnrm	$-0.71(-0.75)$	$-0.58(-0.90)$	$-0.76(-0.66)$
gfdl	$-0.64(-0.79)$	$-0.58(-0.88)$	$-0.84(-0.54)$
ccsm	$-0.78(-0.82)$	$-0.59(-0.83)$	$-1.25(-0.77)$
inm	$-0.81(-0.74)$	$-0.57(-0.85)$	$-1.50(-0.86)$
ipsl	$-0.80(-0.76)$	$-0.43(-0.90)$	$-0.69(-0.64)$
cnf_h	$-0.52(-0.85)$	$-0.51(-0.82)$	$-0.60(-0.92)$
mpi	$-0.96(-0.86)$	$-0.56(-0.79)$	$-0.96(-0.79)$
ukmo	$-0.51(-0.84)$	$-0.66(-0.97)$	$-0.89(-0.92)$
mri	$-0.63(-0.96)$	$-0.77(-0.96)$	$-1.15(-0.88)$

ity and cloud optical thickness under skies with larger cloud fraction. The actual changes are asymptotic with respect to changes in TCF because the cloud optical depth usually tends to increase with increasing cloud fraction.

While a value of -0.71 for the observed $\Delta T_{SW}/\Delta TCF$ is close to that at the SGP and Manus sites, the models give very diverse predictions for $\Delta T_{SW}/\Delta TCF$, ranging from -0.60 in CNF_Hires to -1.50 in INM. Both $\Delta T_{SW}/\Delta TCF$ and TCF (see Fig. 6) are significantly underestimated in CNF_Hires, showing too large transmissivity values for larger TCF in this model. In contrast, $\Delta T_{SW}/\Delta TCF$ is overestimated by more than 100 % in the INM, with an attendant lack of smaller TCF values. Seven models overestimate and two models underestimate the $\Delta T_{SW}/\Delta TCF$ over NSA, with most models exhibiting little stable correlation between TCF and transmissivity, and thus no well-defined bi-modal behavior in the relationship.

5 Evaluation of cloud vertical structure

In Sect. 4 we comprehensively evaluated the TCF simulated by the GCMs and its impact on mean cloud transmissivity and the surface SW flux. Here, we use the long-term ARSCL observations to evaluate the vertical structure of cloud fraction in the models. For simplicity in discussion, low, middle and high clouds are defined as those located at heights of 0–3 km, 3–6 km and higher than 6 km, respectively.

5.1 Manus

Figure 9 (left) shows the annual mean vertical profiles of cloud fraction (CF) derived from ARM ARSCL observations and simulated by seven GCMs over Manus, and the right panel shows the standard deviation among the models. Figure 10 shows the monthly time-height composite plots of CF. Simulated CF in most of the GCMs differs substantially from
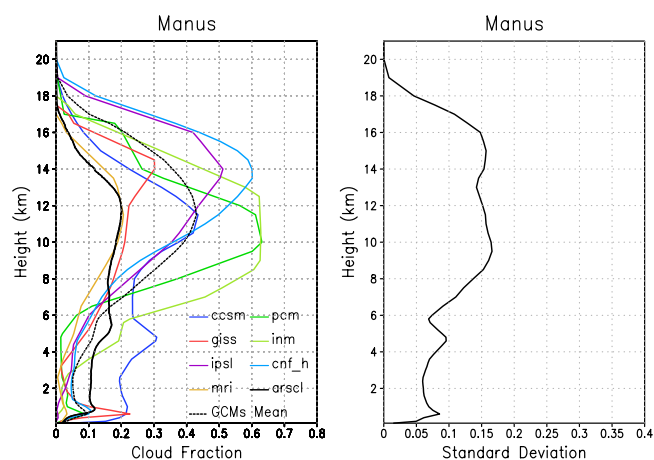
Fig. 9. The annual mean vertical profiles of cloud fraction (CF) from ARSCL observation and GCMs (left) and their standard deviation in seven GCMs (right) over Manus. The GCM results for 1980–1999 and ARSCL observation for 1999–2008 are used in climatology average.

the observed. Overall, most of the GCMs tend to underpredict CF in the lower and mid troposphere. The CCSM is an exception as it significantly overpredicts CF of low and middle clouds. The observation shows a remarkable seasonal variation of CF at different levels (Fig. 10) with a minimum CF in April and a maximum CF in July at lower levels. Except for CCSM, none of the GCMs capture this seasonal variation of CF at lower levels over Manus.

The high level CF in almost all GCMs is substantially larger than the observation over Manus. The high level CF averaged in all 7 GCMs is around 3 times as large as that in the ARSCL observation. This is possibly to some extent a result of the different thresholds determining thin cloud in climate models and ARM measurements, and of the sensitivity limits of the ARM MMCR and MPL for high altitude clouds. The choice of thresholds in determining thin high-cloud is somewhat arbitrary in climate models and in lidar/radar retrievals. The CF at high altitudes is almost linearly dependent on the cutoff value of the optical thickness of these thin high-clouds. Except for the CCSM, none of the GCMs capture the seasonal variability of CF at high level. The GCMs such as INM and CNF_Hires show no seasonal variation for the high level cloud.

Cloud top in most of the GCMs is notably higher in comparison to the ARSCL observations at Manus. The cloud top height is around 17 km in ARSCL, while except for MRI, the cloud top height reaches 19–20 km in most of the GCMs. The cloud top height in ARSCL is probably underestimated to some extent as the radar cannot detect small particles at the top of ice clouds and the lidar is often attenuated in optically thick ice cloud before reaching cloud top. It is interesting that only CCSM can capture the seasonal variability of CF at both low and high levels, although CF in CCSM is larger than
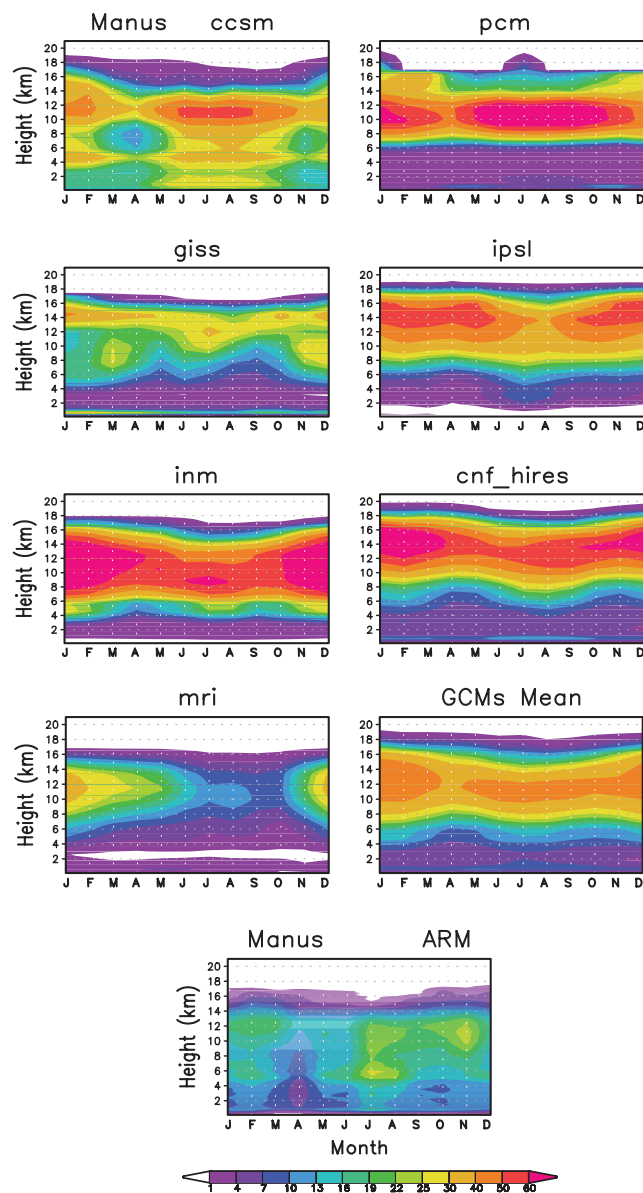

Fig. 10. The monthly mean time-height plots of cloud fraction (CF) from ARSCL observation and GCMs over Manus.

the observation at all levels. Meanwhile, the TCF simulated in CCSM is close to that in ARSCL (Tables 2 and 3), which indicates that the cloud overlap scheme in CCSM produces a result similar to the observation. However, we should keep in mind that the overlap derived from the ARSCL observations is not necessarily the true overlap due to limitations in the measurements (such as the difficulty of lidar to penetrate through thick clouds and the lack of detection of thin layers by radar). The model-measurement difference reaches a maximum at around 8–16 km, which is also the height where the maximum of inter-model deviation of CF is located. Figure 9b shows that the standard deviation (SD) of CF among the 7 GCMs is larger than 0.15 between 8–16 km, while it
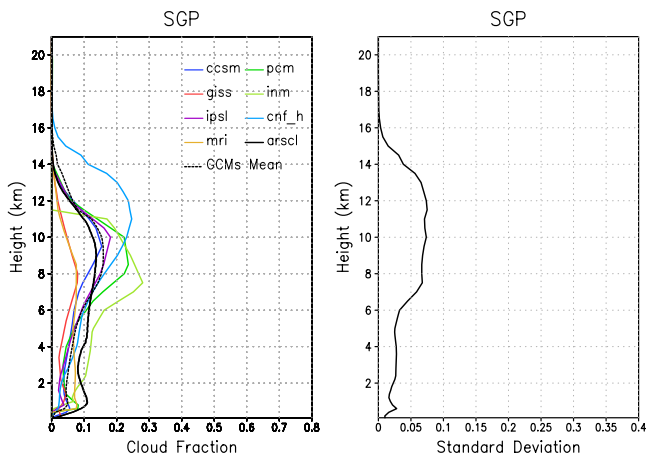
Fig. 11. Same as for Fig. 9, but for SGP.

is less than 0.1 below 6 km. The larger inter-model deviation at higher levels implies the current GCMs have more problems in simulating the high clouds (e.g. cirrus or ice clouds) over the tropical region. Much of the high cloud observed over Manus lie in the outflow from deep convection over the Maritime Continent (Mather, 2005), indicating that the GCMs likely have trouble representing the full radiative impact of tropical deep convection systems. As argued by Waliser et al. (2009), the shortcomings in the representation of these clouds impact both the latent and radiative heating processes, and in turn the circulation and the energy and water cycles, leading to errors in weather and climate forecasts and to uncertainties in quantifying cloud feedbacks associated with global change.

5.2 SGP

Figures 11 and 12 show the annual mean vertical profiles and monthly time-height plots of CF from the ARSCL observation and GCMs over SGP. The observed CF has a bimodal vertical distribution with a higher peak around 6-10 km and a lower one below 2 km. The maximum CF of high clouds occurs during the winter and spring (Fig. 12) when baroclinic wave activity is common over the ARM SGP site (Xi et al., 2010). High-cloud fraction also varies somewhat with the tropopause heights by season due to the change in thermal thickness of the atmosphere. CF is relatively smaller during July to September, especially for low clouds, which is consistent with that for TCF as shown in Fig. 6b.

The GISS and MRI simulate the smallest CF at all levels, while CNF_Hires and INM simulate largest CF at higher levels. Most of the GCMs tend to underpredict CF by 50–150 % at low and middle levels. The CF averaged over all GCMs is around half and two-thirds of the values of ARSCL at low and middle levels, respectively. Except for MRI, all other GCMs fail to capture the distinct boundary layer cloud during winter and spring in their simulations. The ARSCL CF at
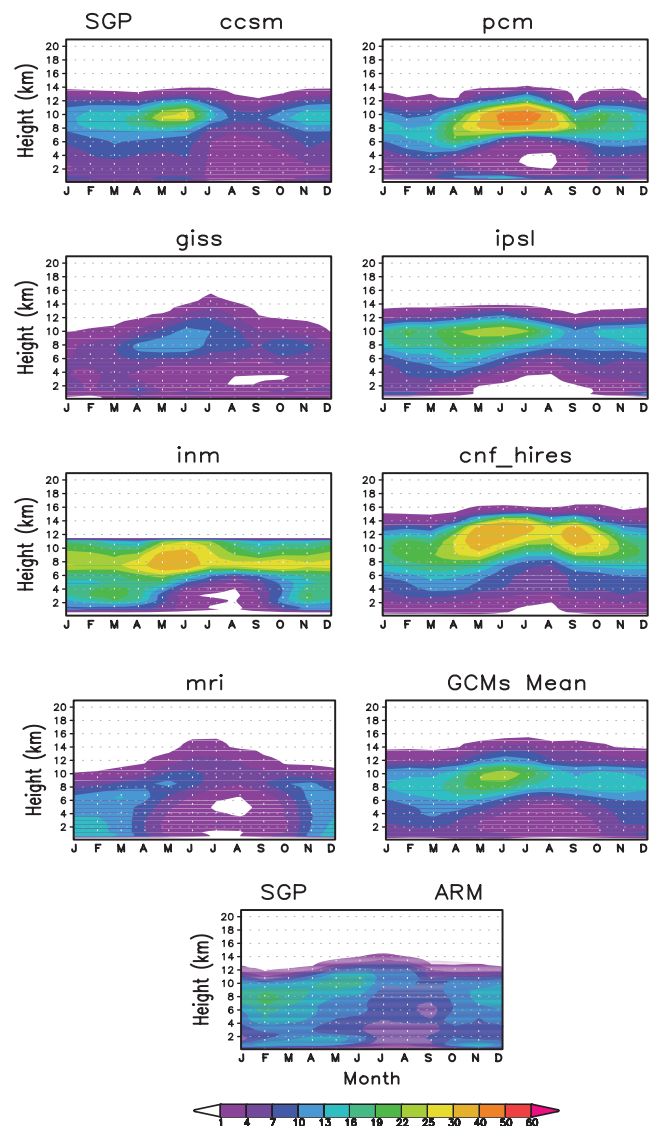

Fig. 12. Same as for Fig. 10, but for SGP.

high level is larger than that predicted in the GISS and MRI, but is smaller than that in the other models. The mean CF of all GCMs is only slightly larger than that for the ARSCL observation at high levels. While most of the models capture the minimum CF at low level during July to September, only CCSM relatively reasonably captures the seasonal variation of CF at high level.

While the model-measurement difference is larger at lower levels below 5 km, the inter-model deviation of CF is larger for high clouds (i.e. above 6 km). Figure 11b shows that the SD of CF among the 7 GCMs is around 0.07 between 7–13 km, and is less than 0.03 below 6 km. The SD for both high and low clouds over SGP is only half as large as that over Manus. The smaller inter-model deviation in SGP suggests that the current GCMs perform more consistently in
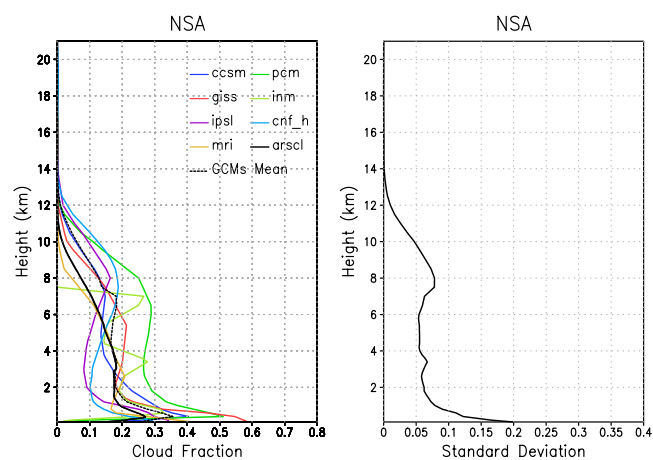
Fig. 13. Same as for Fig. 9, but for NSA.

simulating vertical distribution of CF over the mid-latitude continent than over the tropics, which could partly result from the much stronger large scale forcing for SGP compared to the TWP regime.

5.3 NSA

The observed and simulated annual mean vertical profiles of CF over NSA are shown in Fig. 13. Unlike Manus and SGP, most of the clouds at NSA are low level clouds below 1–2 km. The CF gradually decreases with the height and the maximum cloud top height is around 10–12 km. Therefore, the total cloud cover examined earlier is dominated by low clouds over NSA, either single-layered or multi-layered systems with a significant low-cloud component.

The observed and simulated monthly mean time-height plots of CF are shown in Fig. 14. The maximum CF of low clouds occurs in late spring characterized by optically thin cloud, late-summer and fall with more optically dense clouds. The deeper boundary layer and low clouds in late summer and fall than in late spring is likely due to the retreat of sea ice from the north coast of Alaska, which increases moisture fluxes into the lower atmosphere. In contrast, the cloudiness in May is typical of continental landmasses in spring (i.e. scattered "fair weather" cumulus clouds that form on an otherwise clear day). In a low-solar-zenith-angle environment such as the Alaskan North Slope, the scattered cumulus clouds sideways scatter a significant fraction of the downwelling solar flux to the surface (Dong et al., 2010).

While the GCMs generally capture the maximum CF in the boundary layer and vertical variability (i.e. decreasing with height) of CF, some GCMs (e.g. PCM) tend to overpredict the CF at high level and in the boundary layer. It can be found from Figs. 12 and 14 that a fixed cloud top is probably applied in the INM. Figure 13b shows that the SD of CF among the 7 GCMs is around 0.2 near surface and gradually
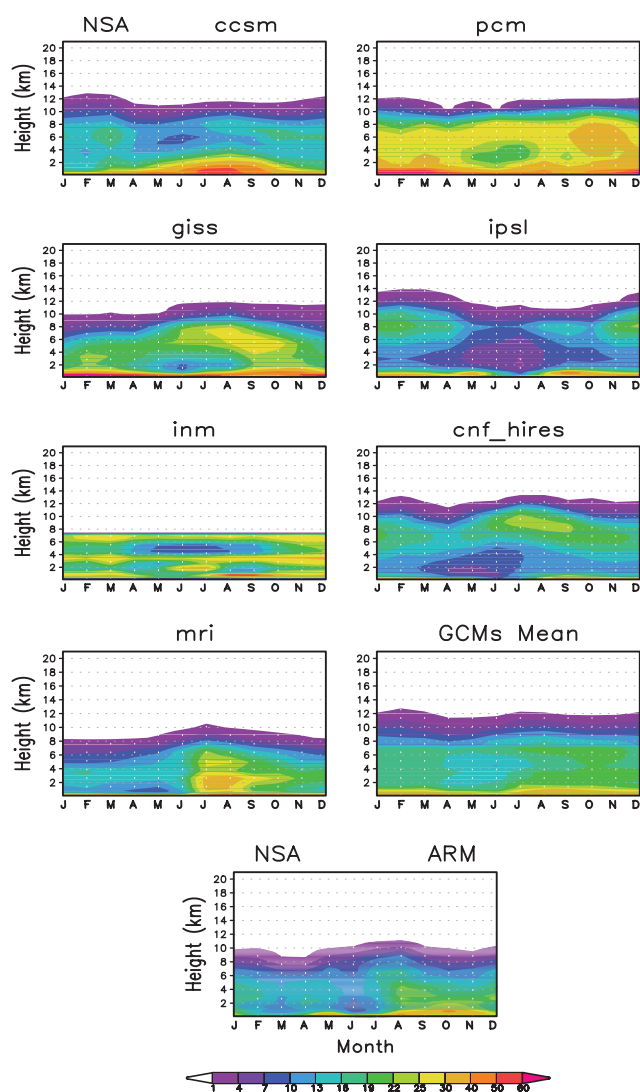

Fig. 14. Same as for Fig. 10, but for NSA.

decreases with height. It becomes constant (around 0.05–0.06) between 2–10 km.

5.4 Variability among ensemble runs within the same GCM

A few IPCC AR4 models have conducted several ensemble simulations, which is important to identify the internal variability and uncertainty of the model results, especially in projecting the future climate change. For example, GISS has four and IPSL has six ensemble simulations in CMIP3. Previous studies have typically focused on evaluating the model internal variability and uncertainty for the simulated temperature and precipitation. It is interesting to examine the internal variability of CF among ensemble simulations.

Figure 15 shows the vertical profiles of CF for four GISS and six IPSL simulations, and their SD among ensemble runs
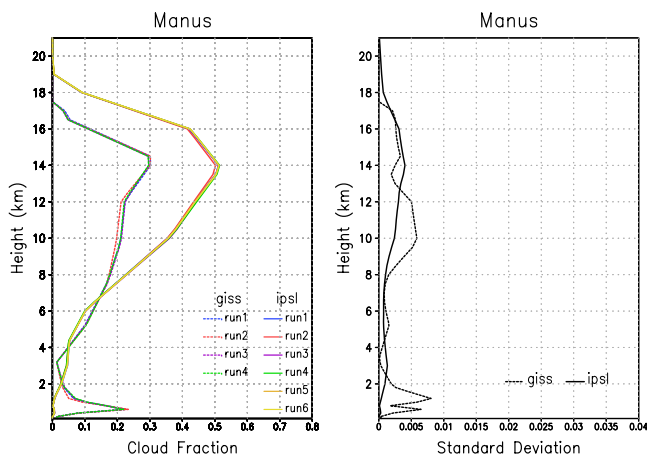
Fig. 15. The vertical profiles of cloud fraction (CF) for four GISS and six IPSL ensemble simulations (left), and their standard deviations (SD, right) among ensemble runs over Manus.

over Manus. The results show that the variability of CF among ensemble runs in the same GCM is minor at all levels. For example, SD for both GISS and IPSL is usually less than 0.005 in Manus, around 2–10 % of inter-model SD. Similar conclusions are found over the SGP and NSA (not shown). This indicates that the internal variability of CF in the same model with ensemble simulations is insignificant.

6   Summary and discussion

Cloud Fraction (CF) has long been recognized as the dominant modulator of radiative fluxes. In this study, we evaluate CF simulations in the IPCC AR4 GCMs against ARM ground measurements at a climatological time-scale, with a focus on the vertical structure, total amount of cloud and its effect on cloud transmissivity, for both inter-model deviation and model-measurement discrepancy. The frequency of hydrometeor occurrence statistics derived from the Active Remotely-Sensed Clouds Locations (ARSCL) observation, the Total Sky Imager (TSI), and the Total Sky Cover (TSK) derived from surface SW radiometers, all are CF or sky cover related products available at ARM sites. Our inter-comparisons reveal that they are correlated with each other but the daily differences are quite significant, suggesting that one should be extremely cautious in using transient (hourly or daily mean) CF data to quantitatively evaluate CF calculated in climate models or retrieved by satellite. A common feature among three sites is that the TSI produces smaller TCF compared to a radar/lidar dataset for highly cloudy days (CF > 0.8), but produces a larger TCF value than the radar/lidar for less cloudy conditions (CF < 0.3). The compensating errors in lower and higher CF days result in small bias of TCF between the vertically pointing radar/lidar dataset and the hemispheric TSI measurements as multi-year

data is averaged. The differences are usually less than 10 % among their multi-year monthly mean values and less than 5 % among their annual mean values, which gives more confidence in using ARSCL CF to evaluate the GCM climatology simulations.

Detailed comparisons of the GCMs results with the ARM observations reveal that the model bias against the observation and the inter-model deviation (disparity) have a similar magnitude for the total CF (TCF) and for the normalized cloud effect, and they are twice as large as that for the surface downward solar radiation and cloud transmissivity. This implies that the other dimensions of cloud, such as cloud optical depth and height, has a similar magnitude of disparity to TCF among the GCMs, and suggests that a better agreement among the GCMs in solar radiative fluxes could be a result from compensating errors in either cloud vertical structure, cloud optical depth or cloud fraction. Similar deviation pattern between inter-model and model-measurement suggests that the climate models tend to generate larger bias against observations for those variables with larger inter-model deviation. The simulated TCF from IPCC AR4 GCMs are very scattered through all seasons over three ARM sites (SGP, Manus and NSA). The GCMs perform better at SGP than at Manus and NSA in simulating the seasonal variation and probability distribution of TCF; however, the TCF in these models is remarkably underpredicted and cloud transmissivity is less susceptible to the change of TCF than the observed at SGP.

Most of the GCMs tend to underpredict CF and fail to capture the seasonal variability of CF at middle and lower levels in the tropics. The high level CF is much higher in the GCMs than the observation and the inter-model variability of CF also reaches maximum at high level in the tropics. Most of the GCMs tend to underpredict the CF by 50–150 % at low and middle levels over SGP. Unlike clouds over Manus and SGP, most of clouds in NSA are in the lower troposphere. While the GCMs generally capture the maximum CF in the boundary layer and vertical variability, the inter-model deviation is largest near the surface over the Arctic. The internal variability of CF simulated in ensemble runs with the same model is very minimal.

While the results in this study could be valuable for advancing our understanding of the CF-related data that are available at ARM sites and for providing insights for the climate modeling community in improving the cloud-radiation interaction and the CF parameterization in climate models, several uncertainties should be taken into account in interpreting the results of this study. The primary one is the cutoff value of the optical depth to define cloudiness in the various observations and climate models. The CF at high altitudes is almost linearly dependent on the cutoff value of the optical thickness of these thin high-clouds. However, the choice of thresholds in determining thin cloud is somewhat arbitrary in climate models and in lidar/radar retrievals. The different cutoff values in defining the clouds

in the ARM measurements and in the GCMs can result in appreciable uncertainty in comparing the CF in the observations and models. This uncertainty will be reduced once the cloud radar/lidar simulators are implemented into GCMs.

The second uncertainty is related to the different spatial coverage between the point observations and model results. The GCM grid variables represent an average in a grid cell (e.g. 200 km × 200 km, which is a typical horizontal grid of AR4 GCMs in lower latitudes). The ARSCL is essentially the vertically pointing instruments with a narrow FOV that can only detect clouds directly above the instruments. This is a common problem when evaluating the model results, although the uncertainty introduced by this factor could be reduced by increasing the model spatial resolution or averaging more years of data.

The third uncertainty is the limited length of the valid measurement data available at the ARM sites. Although the CF and surface radiative fluxes data are available starting from the late 1990s at most of the sites, less than 10 yr of AR-SCL and TSI data are used for some variables in this study because of missing data due to instrument down time. However, the surface radiation data are mostly complete, and we show good agreement between the TSI and TSK TCF values, suggesting that the more continuous TSK data are well suited for long term comparison efforts of TCF. Nevertheless, a 20-yr complete dataset would be more ideal for this kind of study in a climatological sense.

It is highly desirable to see if the positive and negative attributes of model clouds can be associated with specific physical parameterizations. The results of this study show that there is no particular model with a specific cloud scheme that has superior performance in all aspects of CF simulation than other models in all three sites. The underestimation of TCF and the overestimation of optical thickness of clouds are common over SGP to models that used very different cloud schemes, however, this could be due to completely different reasons. As suggested by Webb et al. (2001), many other model components can be as important as the cloud and precipitation schemes in assessing clouds in models, such as the vertical resolution and cloud microphysical properties. Without carrying out controlled experiments by isolating individual physical parameterization components, it is difficult to pinpoint the source of the model differences.

In future work, we plan to do such experiments for the physics parameterizations used in the Community Atmosphere Model (CAM5). Colleagues at PNNL have implemented the CAM5 physics package in the Weather Research and Forecasting (WRF) model (J. Fast, personal communication, 2011), allowing examination of the range of behavior of the physics parameterizations over a range of scales, including those closer to the scale of the ARM observations. The WRF model can be run using all of the CAM5 physics or only individual components, which will allow exploration of the effects of individual parameterizations on the resulting cloud fraction and radiation relationships. The WRF model

can also be run at very high spatial resolution, which can reduce the uncertainty in model evaluation induced by the different spatial coverage between model and measurement. Forcing the model with reanalysis data will also reduce the potential discrepancies in large-scale dynamics between the model and observations that can exist in free-running climate models, and this can be one cause of model/observation disagreement. We will also save hourly output from the model to allow us to investigate the diurnal cycle of cloudiness in parameterization scheme.

Inclusion of a ground-based radar and lidar simulator in the model will allow more direct assessment of cloud overlap, vertical structure, fall velocity, cloud phase, and cloud microphysical assumptions against the ARM radar observations. New radar observations and techniques such as radar spectra measurements provide vertical velocity statistics which can be used to examine assumptions in convective parameterizations (Kollias and Albrecht, 2010), better identification of regions with multi-modal characteristics such as mixed phase regions (Shupe et al., 2004), and better discrimination between cloud and drizzle (Kollias et al., 2011) which will be useful for investigation of autoconversion rates. The satellite simulator has been installed in some of the IPCC AR5 GCMs; we may repeat the analysis for the IPCC AR5 GCMs and compare the results with those from this study.

Edited by: B. Stevens

References

Ackerman, T. P. and Stokes, G. M.: The Atmospheric Radiation Measurement Program, Phys. Today, 56, 38–44, 2003.

Berg, L. and Stull, R.: Accuracy of point and line measures of boundary layer cloud amount, J. Appl. Meteor., 41, 640–650, 2002.

Boers, R., de Haij, M. J., Wauben, W. M. F., Baltink, H. K., van Ulft, L. H., Savenije, M., and Long, C. N.: Optimized fractional cloudiness determination from five ground-based remote sensing techniques, J. Geophys. Res., 115, D24116, doi:10.1029/2010JD014661, 2010.

Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J.-L., Hall, A., Hallegatte, S., Holland, M. M., Ingram, W., Randall, D. A., Soden, D. J., Tselioudis, G., and Webb, M. J.: How Well Do We Understand and Evaluate Climate Change Feedback Processes?, J. Climate, 19, 3445–3482, 2006.

Bretherton, C. S., McCaa, J. R., and Grenier, H.: A new parameterization for shallow cumulus convection and its application to marine subtropical cloud-topped boundary layers. Part I: Description and 1D results, Mon. Weather Rev., 132, 864–882, 2004.

Brooks, M. E., Hogan, R. J., and Illingworth, A. J.: Parameterizing the difference in cloud fraction defined by area and by volume as observed by radar and lidar, J. Atmos. Sci., 62, 2248–2260, 2005.

Cess, R. D., Zhang, M. H., Minnis, P., Corsetti, L., Dutton, E. G., Forgan, B. W., Garber, D. P., Gates, W. L., Hack, J. J., Harrison, E. F., Jing, X., Kiehi, J. T., Long, C. N., Morcrette, J.-J., Potter, G. L., Ramanathan, V., Subasilar, B., Whitlock, C. H., Young, D. F., and Zhou, Y. : Absorption of solar radiation by clouds: Observations versus models, Science, 267, 496–499, 1995.

Chou, M.-D., Suarez, M. J., Ho, C.-H., Yan, M. M.-H., and Lee, K.-T.: Parameterization of cloud overlapping and shortwave single-scattering properties for use in general circulation and cloud ensemble models, J. Climate, 11, 202–214, 1998.

Clothiaux, E. E., Moran, K. P., Martner, B. E., Ackerman, T. P., Mace, G. G., Uttal, T., Mather, J. H., Widener, K. B., Miller, M. A., and Rodriguez, D. J.: The Atmospheric Radiation Measurement Program Cloud Radars: Operational Modes, J. Atmos. Ocean. Tech., 16, 819–827, 1999.

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

Collins, W. D.: Parameterization of generalized cloud overlap for radiative calculations in general circulation models, J. Atmos. Sci., 58, 3224–3242, 2001.

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Kiehl, J. T., Briegleb, B., Bitz, C., Lin, S.-J., Zhang, M., and Dai, Y.: Description of the NCAR Community Atmosphere Model (CAM3.0), Tech. Note TN-464+STR, 214 pp., Natl. Cent. for Atmos. Res. Boulder, Colorado, 2004.

Comstock, J. M., Ackerman, T. P., and Mace, G. G.: Ground-based lidar and radar remote sensing of tropical cirrus clouds at Nauru Island: Cloud statistics and radiative impacts, J. Geophys. Res., 107, 4714, doi:10.1029/2002JD002203, 2002.

Curry, J. A., Rossow, W. B., Randall, D., and Schramm, J. L.: Overview of Arctic cloud and radiation characteristics, J. Clim., 9, 1731–1764, 1996.

Del Genio, A. D., Yao, M. S., Kovari, W., and Lo, K. K. W.: A prognostic cloud water parameterization for global climate models, J. Climate, 9, 270–304, 1996.

Dong, X. Q., Xi, B. K., and Minnis, P.: 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, 2006.

Dong, X., Xi, B., Crosby, K., Long, C. N., Stone, R. S., and Shupe, M. D.: A 10 year climatology of Arctic cloud fraction and radiative forcing at Barrow, Alaska, J. Geophys. Res., 115, D17212, doi:10.1029/2009JD013489, 2010.

Durr, B. and Philipona, R.: Automatic cloud amount detection by surface longwave downward radiation measurements, J. Geophys. Res., 109, D05201, doi:10.1029/2003JD004182, 2004.

Garratt, J. R.: Incoming shortwave fluxes at the surface – a comparison of GCM results with observations, J. Climate, 7, 72–80, 1994.

Geleyn, J.-F. and Hollingsworth, A.: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation, Beitr. Phys. Atmos., 52, 1–16, 1979.

Hahn, C., Rossow, W., and Warren, S.: ISCCP cloud properties associated with standard cloud types identified in individual surface observation, J. Climate, 14, 11–28, 2001.

Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Brooks, M. E., Delano, J., Donovan, D. P., Eastment, J. D., Gaussiat, N., Goddard, J. W. F., Haeffelin, M., Klein Baltink, H., Krasnov, O. A., Pelon, J., Piriou, J.-M., Protat, A., Russchenberg, H. W. J., Seifert, A., Tompkins, A. M., van Zadelhoff, G.-J., Vinit, F., Willén, U., Wilson, D. R., and Wrench, C. L.: Cloudnet – Continuous evaluation of cloud profiles in seven operational models using ground-based observations, B. Am. Meteorol. Soc., 88, 883–895, 2007.

Kassianov, E., Long, C. N., and Ovtchinnikov, M.: Cloud Sky Cover versus Cloud Fraction: Whole-Sky Simulations and Observations, J. Appl. Meteor., 44, 86–98, doi:10.1175/JAM-2184.1, 2005.

Kennedy, A. D., Dong, X., Xi, B., Minnis, P., Del Genio, A. D., Wolf, A. B., and Khaiyer, M. M.: 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, 2010.

Kollias, P. and Albrecht, B.: Vertical velocity statistics in fair weather cumuli at the ARM TWP Nauru climate research facility, J. Clim., 23, 6590–6604, doi:10.1175/2010JCLI3449.1, 2010.

Kollias, P., Remillard, J., Luke, E., and Szyrmer, W.: Cloud radar Doppler spectra in drizzling stratiform clouds: 1. Forward modeling and remote sensing applications, J. Geophys. Res., 116, D13201, doi:10.1029/2010JD015237, 2011.

Li, Z., Moreau, L., and Arking, A.: On solar energy disposition: a perspective from observation and modeling, B. Am. Meteorol. Soc., 78, 53–70, 1997.

Liu, Y., Wu, W., Jensen, M. P., and Toto, T.: Relationship between cloud radiative forcing, cloud fraction and cloud albedo, and new surface-based approach for determining cloud albedo, Atmos. Chem. Phys., 11, 7155–7170, doi:10.5194/acp-11-7155-2011, 2011.

Long, C. N.: Correcting for Circumsolar and Near-Horizon Errors in Sky Cover Retrievals from Sky Images, The Open Atmospheric Science Journal, 4, 45–52, doi:10.2174/1874282301004010045, 2010.

Long, C. N. and Shi, Y.: The QCRad Value Added Product: Surface Radiation Measurement Quality Control Testing, Including Climatologically Configurable Limits, Atmospheric Radiation Measurement Program Technical Report, ARM TR-074, 69 pp., available at: http://www.arm.gov, 2006.

Long, C. N. and Shi, Y.: An automated quality assessment and control algorithm for surface radiation measurements, Open Atmos. Sci. J., 2, 23–37, doi:10.2174/1874282300802010023, 2008.

Long, C. N., Slater, D. W., and Tooman, T.: Total Sky Imager (TSI) Model 880 Status and Testing Results, Atmospheric Radiation Measurement Program Technical Report, ARM TR-006, 36 pp., available at: http://www.arm.gov, 2001.

Long C. N., Ackerman, T. P., Gaustad, K. L., and Cole, J. N.: Estimation of Fractional Sky Cover from Broadband Shortwave Radiometer Measurements, J. Geophys. Res.-Atmos., 111, D11204, doi:10.1029/2005JD006475, 2006.

Long, C. N., Dutton, E. G., Augustine, J. A., Wiscombe, W., Wild, M., McFarlane, S. A., and Flynn, C. J.: Significant decadal brightening of downwelling shortwave in the continental United States, J. Geophys. Res., 114, D00D06, doi:10.1029/2008JD011263, 2009.

Mace, G. G. and Benson, S.: The vertical structure of cloud occurrence and radiative forcing at the SGP ARM site as revealed by 8 years of continuous data, J. Climate, 21, 2591–2610, 2008.

Marchand, R., Haynes, J., and Mace, G. G.: A comparison of simulated cloud radar output from the multiscale modeling framework global climate model with CloudSat cloud radar observations, J. Geophys. Res., 114, D00A20, doi:10.1029/2008JD009790, 2009.

Mather, J. H.: Seasonal variability in clouds and radiation at the Manus ARM Site, J. Climate, 18, 2417–2428, 2005.

McFarlane, S. A. and Evans, K. F.: Clouds and shortwave fluxes at Nauru. Part II: Shortwave flux closure, J. Atmos. Sci., 61, 2602–2615, 2004.

McFarlane, S. A., Mather, J. H., and Ackerman, T. P.: Analysis of tropical radiative heating profiles: A comparison of models and observations, J. Geophys. Res., 112, D14218, doi:10.1029/2006JD008290, 2007.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., Stouffer, R. J., and Taylor, K. E.: The WCRP CMIP3 multi-model dataset: A new era in climate change research, B. Am. Meteorol. Soc., 88, 1383–1394, 2007.

Naud, C. M., Del Genio, A., Mace, G. G., Benson, S., Clothiaux, E. E., and Kollias, P.: Impact of Dynamics and Atmospheric State on Cloud Vertical Overlap, J. Clim., 21, 1758–1770, doi:10.1175/2007JCLI1828.1, 2008.

Potter, G. L. and Cess, R. D.: Testing the impact of clouds on the radiation budgets of 19 atmospheric general circulation models, J. Geophys. Res., 109, D02106, doi:10.1029/2003JD004018, 2004.

Randall, D., Coakley Jr., J., Fairall, C., Kropfli, R., and Lenschow, D.: Outlook for research on subtropical marine stratiform clouds, B. Am. Meteor. Soc., 65, 1290–1301, 1984.

Randall, D., Cess, R. D., Blanchet, J. P., Boer, G. J., Dazlich, D. A., Del Genio, A. D., Deque, M., Dymnikov, V., Galin, V., Ghan, S. J., Lacis, A. A., Le Treut, H., Li, Z.-X., Liang, X.-Z., McAvaney,

B. J., Meleshko, V. P., Mitchell, J. F. B., Morcrette, J.-J., Potter, G. L., Rikus, L., Roeckner, E., Royer, J. F., Schlese, U., Sheinin, D. A., Slingo, J., Sokolov, A. P., Taylor, K. E., Washington, W. M., Wetherald, R. T., Yagai, I., and Zhang, M.-H.: Intercomparison and interpretation of surface energy fluxes in atmospheric general circulation models, J. Geophys. Res., 97, 3711–3725, 1992.

Rossow, W. B. and Schiffer, R. A.: ISCCP cloud data products, B. Am. Meteor. Soc., 72, 2–20, 1991.

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'Connor, E. J., Rossow, W. B., Durden, S. L., Miller, S. D., Austin, R. T., Benedetti, A., Mitrescu, C., and the CloudSat Science Team: : The cloudsat mission and the A-train, B. Am. Meteorol. Soc., 83, 1771–1790, doi:10.1175/BAMS-83-12-1771, 2002.

Shi, Y. and Long, C. N.: Best Estimate Radiation Flux Value Added Product: Algorithm Operational Details and Explanations, Atmospheric Radiation Measurement Program Technical Report, ARM TR-008, 58 pp., available at: http://www.arm.gov, 2002.

Shupe, M., Kollias, P., Matrosov, S., and Schneider, T.: Deriving Mixed-Phase Cloud Properties from Doppler Radar Spectra, J. Atmos. Tech., 21, 660–670, 2004.

Stone, R. S., Dutton, E. G., Harris, J. M., and Longenecker, D.: Earlier spring snowmelt in northern Alaska as an indicator of climate change, J. Geophys. Res., 107, 4089, doi:10.1029/2000JD000286, 2002.

Taylor, K. E.: Climate Radiative Responses and Clouds: What we are learning from CMIP3, 18TH ARM Science Team Meeting Norfolk, VA 13 March 2008.

Tian, L. and Curry, J. A.: Cloud overlap statistics, J. Geophys. Res., 94, 9925–9935, 1989.

Vavrus, S., Waliser, D., Schweiger, A., and Francis, J.: Simulations of 20th and 21st century Arctic cloud amount in the global climate models assessed in the IPCC AR4, Clim. Dynam., 33, 1099–1115, doi:10.1007/s00382-008-0475-6, 2009.

Waliser, D. E., Li, J.-L. F., Woods, C. P., Austin, R. T., Bacmeister, J., Chern, J., Genio, A. D., Jiang, J. H., Kuang, Z., Meng, H., Minnis, P., Platnick, S., Rossow, W. B., Stephens, G. L., Sun-Mack, S., Tao, W.-K., Tompkins, A. M., Vane, D. G., Walker, C., and Wu, D.: Cloud ice: A climate model challenge with signs and expectations of progress, J. Geophys. Res., 114, D00A21, doi:10.1029/2008JD010015, 2009.

Walsh, J. E., Chapman, W. L., and Portis, D. H.: Arctic Cloud Fraction and Radiative Fluxes in Atmospheric Reanalyses, J. Climate, 22, 2316–2334, doi:10.1175/2008JCLI2213.1, 2009.

Wang, Y., Long, C. N., Leung, L. Y. R., Dudhia, J., McFarlane, S. A., Mather, J. H., Ghan, S. J., and Liu, X.: Evaluating regional cloud-permitting simulations of the WRF model for the Tropical Warm Pool International Cloud Experiment (TWP-ICE, Darwin 2006), J. Geophys. Res.-Atmos., 114, D21203, doi:10.1029/2009JD012729, 2009.

Warren, S. G., Eastman, R. M., and Hahn, C. J.: A survey of changes in cloud cover and cloud types over land from surface observations, 1971–96, J. Climate, 20, 717–738, 2007.

Webb, M., Senior, C., Bony, S., and Morcrette, J.-J.: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models, Clim. Dynam., 17, 905–922, 2001.

Wild, M.: Solar radiation budgets in atmospheric model intercom-

parisons from a surface perspective, Geophys. Res. Lett., 32, L07704, doi:10.1029/2005GL022421, 2005.

Wild, M.: Short-wave and long-wave surface radiation budgets in GCMs: a review based on the IPCC-AR4/CMIP3 models, Tellus A, 60A, 932–945, doi:10.1111/j.1600-0870.2008.00342.x, 2008.

Wild, M., Ohmura, A., Gilgen, H., and Roeckner, E.: Validation of GCM simulated radiative fluxes using surface observations, J. Climate, 8, 1309–1324, 1995.

Wild, M., Ohmura, A., Gilgen, H., Roeckner, E., Giorgetta, M., and Morcrette, J. J.: The disposition of radiative energy in the global climate system: GCM versus observational estimates, Clim. Dy-nam., 14, 853–869, 1998.

Winker, D. M., Pelon, J., and McCormick, M. P.: The CALIPSO mission: Spaceborne lidar for observation of aerosols and clouds, Proc. SPIE, 4893, 1–11, 2003.

Wu, D. L., Ackerman, S. A., Davies, R., Diner, D. J., Garay, M. J., Kahn, B. H., Maddux, B. C., Moroney, C. M., Stephens, G. L., Veefkind, J. P., and Vaughan, M. A.: Vertical distributions and re-lationships of cloud occurrence frequency as observed by MISR, AIRS, MODIS, OMI, CALIPSO, and CloudSat, Geophys. Res. Lett., 36, L09821, doi:10.1029/2009GL037464, 2009.

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

Xie S., McCoy, R., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Clothiaux, E. E., Gaustad, K. L., Golaz, J. C., Hall, S., Jensen, M., Johnson, K. L., Lin, Y., Long, C. N., Mather, J. H., McCord, R. A., McFarlane, S. A., Palanisamy, G., Shi, Y., and Turner, D. D.: ARM Climate Modeling Best Estimate Data – A new data product for climate modelers, B. Am. Meteorol. Soc., 91, 13–20, doi:10.1175/2009BAMS2891.1, 2010

Zdunkowski, W. G., Panhans, W.-G., Welch, R. M., and Korb, G. J.: A radiation scheme for circulation and climate models, Contrib. Atmos. Phys., 55, 215–238, 1982.

Zhang, M. H., Lin, W. Y., Klein, S. A., Bacmeister, J. T., Bony, S., Cederwall, R. T., Del Genio, A. D., Hack, J. J., Loeb, N. G., Lohmann, U., Minnis, P., Musat, I., Pincus, R., Stier, P., Suarez, M. J., Webb, M. J., Wu, J. B., Xie, S. C., Yao, M. S., and Zhang, J. H.: Comparing clouds and their seasonal variations in 10 atmo-spheric general circulation models with satellite measurements, J. Geophys. Res., 110, D15S02, doi:10.1029/2004JD005021, 2005.