

Exposing Global Cloud Biases in the Community Atmosphere Model (CAM) Using Satellite Observations and Their Corresponding Instrument Simulators

J. E. KAY,* B. R. HILLMAN,^{+,#} S. A. KLEIN,[@] Y. ZHANG,[@] B. MEDEIROS,* R. PINCUS,[&]
A. GETTELMAN,* B. EATON,* J. BOYLE,[@] R. MARCHAND,⁺ AND T. P. ACKERMAN^{+,#}

* *Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado*

⁺ *Department of Atmospheric Sciences, University of Washington, Seattle, Washington*

[#] *Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington*

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

[&] *NOAA/Earth System Research Lab, Physical Sciences Division, and University of Colorado, Boulder, Colorado*

(Manuscript received 24 August 2011, in final form 1 February 2012)

ABSTRACT

Satellite observations and their corresponding instrument simulators are used to document global cloud biases in the Community Atmosphere Model (CAM) versions 4 and 5. The model–observation comparisons show that, despite having nearly identical cloud radiative forcing, CAM5 has a much more realistic representation of cloud properties than CAM4. In particular, CAM5 exhibits substantial improvement in three long-standing climate model cloud biases: 1) the underestimation of total cloud, 2) the overestimation of optically thick cloud, and 3) the underestimation of midlevel cloud. While the increased total cloud and decreased optically thick cloud in CAM5 result from improved physical process representation, the increased midlevel cloud in CAM5 results from the addition of radiatively active snow. Despite these improvements, both CAM versions have cloud deficiencies. Of particular concern, both models exhibit large but differing biases in the subtropical marine boundary layer cloud regimes that are known to explain intermodel differences in cloud feedbacks and climate sensitivity. More generally, this study demonstrates that simulator-facilitated evaluation of cloud properties, such as amount by vertical level and optical depth, can robustly expose large and at times radiatively compensating climate model cloud biases.

1. Motivation

a. Using satellite simulators to evaluate climate model clouds

Cloud feedbacks dominate uncertainty in model climate projections (e.g., Cess et al. 1990; Bony and Dufresne 2005; Williams and Webb 2009; Medeiros et al. 2008), but the quantification of model cloud biases is often confounded by poor model–observational comparison techniques. In the last decade, data from a number of new cloud-observing satellite platforms have become available. Given this context, the use of satellite simulators to evaluate climate model clouds is an exciting, and unsurprisingly burgeoning research area.

Development of satellite simulation software for evaluation and intercomparison of climate model clouds

was first completed for the International Satellite Cloud Climatology Project (ISCCP). The ISCCP observations (Rossow and Schiffer 1999) and corresponding ISCCP simulator (Klein and Jakob 1999; Webb et al. 2001) have now been applied for over a decade (e.g., Norris and Weaver 2001; Lin and Zhang 2004; Zhang et al. 2005, hereafter Z05; Schmidt et al. 2006; Cole et al. 2011). Common climate model biases revealed by these ISCCP comparison studies include 1) underestimation of total cloud, 2) overestimation of optically thick cloud, and 3) underestimation of midtopped cloud.

More recently, the Cloud Feedbacks Model Intercomparison Project (CFMIP) (Bony et al. 2011) has been coordinating development of the CFMIP Observation Simulator Package (COSP). The abundance of new satellite observations and corresponding diagnostics available in COSP provides new opportunities to understand and quantify climate model cloud biases. As described in Bodas-Salcedo et al. (2011, hereafter B11), COSP currently produces climate model diagnostics that can be

Corresponding author address: J. E. Kay, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.
E-mail: jenkay@ucar.edu

TABLE 1. Comparison of cloud-related physical parameterizations in CAM5 and CAM4.

Parameterization	CAM5 (Neale et al. 2011a), released in June 2010, 30 vertical levels	CAM4 (Neale et al. 2011b), released in April 2010, 26 vertical levels
Radiation	Rapid Radiative Transfer Model Global (RRTMG) (Iacono et al. 2008)	Collins et al. 2001
Shallow convection	Mass flux scheme with convective inhibition (CIN) closure (Park and Bretherton 2009)	Three-level adjustment of moist static energy (Hack 1994)
Deep convection	Bulk mass flux with CAPE closure (Neale et al. 2008)	Bulk mass flux with CAPE closure (Neale et al. 2008)
Planetary boundary layer and turbulence	Moist turbulence scheme based on diagnostic TKE (Bretherton and Park 2009)	Dry turbulence scheme based on specified K profile (Holtslag and Boville 1993)
Cloud microphysics and macrophysics	Prognostic double moment microphysics (Morrison and Gettelman 2008) with ice supersaturation (Gettelman et al. 2010), diagnostic precipitation at each model level, and diagnostic cloud fraction scheme	Prognostic single-moment microphysics, diagnostic precipitation at surface, and diagnostic cloud fraction (Rasch and Kristjansson 1998)
Aerosols	Modal aerosol model (Liu et al. 2011)	Bulk aerosol model

compared to observations from six satellite projects: 1) ISCCP, 2) Multiangle Imaging Spectroradiometer (MISR), 3) Moderate Resolution Imaging Spectroradiometer (MODIS), 4) *CloudSat*—a spaceborne radar, 5) a spaceborne lidar on the Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO), and 6) Polarization and Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a lidar (PARASOL). Although ISCCP observations are available over the longest period (1983–present) of any of the satellite datasets, MODIS (2002–present), and MISR (2000–present) observations are derived from more sophisticated passive retrievals based on more angles (MISR) and more spectral bands (MODIS). Though their observational records are short (2006–present), the active satellite measurements from the spaceborne radar *CloudSat* and the spaceborne lidar CALIPSO measure cloud height directly and offer a distinct advantage over passive instruments for evaluation of cloud vertical structure in climate models (e.g., Chepfer et al. 2008, hereafter C08; Bodas-Salcedo et al. 2008, hereafter B08); Zhang et al. 2010). *CloudSat* radar reflectivities also provide the best precipitation diagnostic that is currently available in COSP.

b. Study goals

The primary goal of this study is to evaluate mean state clouds in the Community Atmosphere Model (CAM) using multiple independent satellite datasets and their corresponding instrument simulators in COSP. The main findings expose large, and at times compensating, model cloud biases both globally and in key climatic regions. The presented model cloud biases could result from both model circulation (dynamics) and cloud parameterization (physics) errors, and it is

beyond the scope of this study to identify their specific causes. Simulator and observational error are thought to be less important for explaining many of the identified model biases. A secondary goal of this study is to summarize robust information gained from the large number of new COSP diagnostics. Previous model evaluation papers using instrument simulators have focused either on diagnostics from a small subset of COSP simulators (e.g., Z05; C08; B08) or on a single geographic region (e.g., B11). Section 2 describes our methodology including a description of the COSP implementation in CAM, the CAM runs analyzed, and the cloud evaluation strategy. Section 3 contains the results. We identify compensating errors between cloud fraction and cloud optical properties that are not apparent from the radiative fluxes or cloud forcing alone. We also find that, while it has deficiencies, the most recent CAM version, CAM5 (Neale et al. 2011a), has significantly reduced three long-standing climate model cloud biases. Finally, we evaluate CAM clouds in select climatically important regions and discuss some limitations of COSP for climate model evaluation. Sections 4 and 5 contain a discussion of uncertainty and the summary.

2. Methods

a. COSP integration into CESM

The Community Earth System Model version 1 (CESM1) includes two CAM physics packages: CAM4 (Neale et al. 2011b) and CAM5 (Neale et al. 2011a). Although both CAM versions are viable options within the CESM1 modeling framework, their only shared physical parameterization is the deep convection scheme (Table 1). This work uses CAM5.1, which is a part of CESM1.0.3, and can run COSP v1.3 inline and produce COSP diagnostics

with either CAM4 or CAM5 physics. COSP-enabled CAM code is available to the general scientific community as a part of CESM public releases (code available to download at <http://www.cesm.ucar.edu/models/cesm1.0/>).

Local modifications to the COSP v1.3 code were necessary both to ensure compatibility with the CESM code and software engineering requirements and to incorporate the influence of radiatively active snow, a modification that only influences the ISCCP, MISR, MODIS, and lidar diagnostics in the CAM5 simulations. Snow in this context represents the population of large ice crystals with appreciable fall velocities. Because it incorporates the impact of snow on radiative fluxes, CAM5 is atypical and is more consistent with observations (e.g., Hogan et al. 2001; Waliser et al. 2011).

Incorporating snow into all COSP simulators is consistent with the notion that simulators should diagnose all hydrometeors that contribute to the signal satellites observe. Within the CESM version of COSP, the subcolumns passed to each individual instrument simulator can contain water clouds, ice clouds, rain, and snow. The distribution of precipitation within each subcolumn is based on the clouds in each subcolumn, as described in Zhang et al. (2010). Code modifications to COSP were made such that subcolumns containing stratiform snow contribute to the ISCCP, MISR, and MODIS cloud diagnostics using the CAM5-predicted shortwave snow optical depth and longwave snow emissivity. Similarly, subcolumns containing stratiform snow contribute to the extinction of the simulated lidar beam using a snow mixing ratio deduced from the snow precipitation flux. Within the lidar simulator, snow is assumed to have the same backscatter-to-extinction ratio as is used for cloud ice (Chepfer et al. 2007). No modifications to the COSP radar simulator were made because snow already contributed to radar reflectivities.

b. CAM simulations with COSP

To enable intercomparison, both atmospheric model physics packages (CAM4 and CAM5) were run with the same finite volume dynamical core, the same prognostic land model [Community Land Model version 4 (CLM4); Lawrence et al. (2011)], and the same observed monthly evolving sea surface temperature and sea ice boundary conditions (Hurrell et al. 2008). Both model physics had tuning parameters set to be consistent with CESM contributions to the Coupled Model Intercomparison Project 5 (CMIP5) (Taylor et al. 2012). In all simulations, COSP diagnostic model cloud fields were produced every 3 h. The active simulators (*CloudSat* radar, CALIPSO lidar) were run on all columns, while the passive simulators (ISCCP, MODIS, MISR) were run on sunlit columns.

Unless otherwise stated, our analysis focuses on 10-yr CAM4 and CAM5 simulations (January 2001–December 2010) run on the standard $0.9^\circ \times 1.25^\circ$ horizontal grid. To evaluate the influence of radiatively active snow on COSP diagnostics, an additional 3-yr CAM5 simulation (January 2001–December 2003) was run using the same $0.9^\circ \times 1.25^\circ$ horizontal resolution, but neglecting all snow-related COSP inputs. Finally, to assess the impact of model horizontal resolution, we ran 5-yr CAM4 and CAM5 simulations (January 2001–December 2005) on a $1.9^\circ \times 2.5^\circ$ horizontal grid.

A number of observations were used to evaluate the mean climate of these COSP-enabled simulations. We compared CAM fluxes to observed top-of-atmosphere (TOA) fluxes in version 2.6 of the Clouds and the Earth's Radiant Energy System (CERES)–Energy Balanced and Filled (EBAF) dataset (Loeb et al. 2009), a dataset currently available over a 10-yr period from March 2000 to February 2010. To evaluate COSP-diagnosed cloud fields, we used observations designed for these comparisons (available at <http://climserv.ipsl.polytechnique.fr/fr/cfmip-observations.html>). As described in B11, these observations include 1) a CALIPSO–GCM Oriented Cloud CALIPSO Product (GOCCP) climatology over the 5-yr period June 2006–November 2010 (Chepfer et al. 2010), 2) a *CloudSat* climatology over the 5-yr period June 2006–November 2010, 3) an ISCCP climatology over the 25-yr period July 1983–June 2008 for daytime only (Pincus et al. 2012, hereafter P12), 4) a MODIS climatology (based on both *Aqua* and *Terra*) over the 8-yr period July 2002–July 2010 for daytime only (P12), and 5) a MISR climatology over the 10-yr period March 2000 to November 2009 for daytime and over oceans only (Marchand et al. 2010, hereafter M10).

c. COSP diagnostics available for global climate model evaluation

COSP outputs include a variety of cloud property and cloud fraction diagnostics that enable consistent inter-model and observation–model comparisons. For example, the MODIS simulator produces liquid effective radius and cloud water path diagnostics, while the ISCCP simulator estimates cloud albedo. Of particular interest in the suite of COSP diagnostics are two-dimensional histograms of cloud fraction as a function of height (or pressure) and a cloud property. These histograms are directly analogous to satellite products that have been popularized over the last decade (e.g., Tselioudis et al. 2000; Jakob and Tselioudis 2003; Rossow et al. 2005; M10). For the three passive instruments, the cloud property is a column-integrated visible optical depth (τ), and the diagnostic is a joint histogram of the highest retrieved cloud-top height (CTH) in the column and τ , herein called a CTH- τ

histogram. For *Cloudsat* and CALIPSO, the histograms are a function of cloud height and radar reflectivity and lidar scattering ratio (SR), respectively. Unlike CTH- τ histograms, the *CloudSat* and CALIPSO two-dimensional histograms cannot be summed to obtain the total cloud fraction because the active instruments measure cloud height not cloud-top height. Also unlike CTH- τ histograms, the cloud property is the observed radar reflectivity or lidar SR and therefore must be related to cloud or precipitation physical characteristics using expert knowledge.

Although challenging, efforts to independently assess uncertainty in satellite retrievals and instrument simulators are vital. For example, Mace et al. (2011, hereafter M11) found more thin ($\tau < 1.3$) cloud in ISCCP satellite observations than in ground-based ISCCP-like observations at the Atmospheric and Radiation Measurement (ARM) Southern Great Plains (SGP) site in Oklahoma. Above $\tau = 1.3$, M11 found that the ground-based ISCCP-like observations had a flatter τ distribution than ISCCP satellite observations. These differences are a reminder of the uncertainties in simulator-based comparisons; however, because the τ distributions at the SGP grid box in CAM differ substantially from the global CAM τ distribution (not shown), there is some question about the global representativeness of validation results from a single midlatitude land site. In addition, our experience shows that differences between observed and CAM cloud τ distributions are often much larger than simulator comparison technique uncertainties.

Despite their limitations, the diagnostics produced by COSP revolutionize comparisons between satellite observations and climate models. For example, the minimum relative humidity to form cloud and other parameters in the CAM physics parameterizations are commonly tuned both to achieve the global radiative balance required for stable climate simulations and to offset errors in the predicted cloud optical depth and vertical distribution. Because of the carefully defined comparisons they enable, COSP diagnostics can be used to evaluate these tuning efforts with multiple independent observational datasets.

To illustrate the importance of simulators for both observation–model comparisons and model intercomparisons, we describe a particularly striking example. Owing to an inconsistency in the treatment of cloud fraction and cloud liquid water, CAM4 can produce stratocumulus cloud with no associated water content and, thus, no impact on radiative fluxes (see also Hannay et al. 2009; Medeiros et al. 2012). As a result of these so-called “empty” clouds, use of the standard model diagnostic cloud fraction in stratocumulus regions is misleading. If one were to use the standard model

diagnostic cloud fraction when comparing CAM4 to another model, or CAM4 to observations, empty clouds would artificially inflate the CAM4 low cloud fractions, in some cases by up to 0.25. A simulator-diagnosed cloud fraction does not suffer from this error because the simulators ignore empty clouds and only count the clouds with sufficient radiative impact to be above the cloud-detection limits of the satellite.

d. A strategy for climate model evaluation using COSP diagnostics

Understanding the utility of the large number of newly available COSP diagnostics for climate model development and evaluation is just beginning. Our strategy is to leverage the strengths of each observational dataset and, where multiple datasets provide a robust and consistent piece of information, to use the collection as a measure of observational uncertainty. Here, we briefly highlight previous studies that influenced our choice of COSP diagnostics to emphasize in our evaluations.

For the passive instruments (MISR, MODIS, ISCCP), M10 and P12 guided our evaluation strategy. M10 compare and discuss differences in CTH- τ observations for all three passive instruments, while P12 discuss the differing philosophies and resulting artifacts in the ISCCP and MODIS simulators and observations. Under strong inversions (e.g., subtropical subsidence cloud regimes off western coasts), temperature inversions confound MODIS and ISCCP cloud-top height retrievals resulting in negative cloud-top pressure (positive cloud-top height) biases of up to 50–200 mb (1–2 km) (M10). As such, the heights of low clouds under inversions are more reliably retrieved from MISR and especially CALIPSO. M10 and P12 both describe detection and retrieval application differences in broken cloud regimes, such as trade cumulus, and in multilayer cloud regimes. In trade cumulus regimes, ISCCP fails to detect many small cumuli and underestimates cloud fraction by as much as 0.10 relative to MISR (M10). MODIS cloud property retrievals are often not performed on cloud edges, resulting in a global annual-mean total-detected cloud fraction that is 0.15 lower than that of ISCCP or MISR (P12). The MODIS cloud deficit mostly comprises low clouds with $\tau < 1.3$ in the MISR and ISCCP dataset (P12). When optically thin cirrus overlay low clouds, ISCCP erroneously reports a midtopped cloud and therefore overestimates midtopped cloud when compared to MODIS and MISR (M10). In contrast, MISR reports the height of the low cloud deck and therefore underreports high cloud amounts when compared to ISCCP and MODIS (M10). P12 found general agreement in the frequency of occurrence of clouds as a function of τ (τ distributions) provided

that $\tau > 1.3$. Based on the above, we emphasize MODIS results for passive high-topped cloud comparisons, MISR results for passive low-topped cloud comparisons but utilize all three passive instruments equally when evaluating τ distributions of all clouds for $\tau > 1.3$.

For the active instruments (*CloudSat*, CALIPSO), studies such as those by B08 and C08 influence our evaluation strategy. B08 used *CloudSat* observations to evaluate the Met Office (UKMO) weather forecast model, and their findings motivated our midlevel cloud and precipitation bias evaluation using *CloudSat*. B08 showed that the UKMO model lacks midlevel *CloudSat* clouds, reaffirming a common climate model bias found in ISCCP-based studies. C08 evaluated climate model clouds using the lidar simulator and CALIPSO observations. While clouds frequently attenuate the lidar beam, C08 report the main strength of using the lidar is that it enables accurate evaluation of cloud vertical distribution. Though not discussed in C08, CALIPSO provides the best currently available satellite observations of polar clouds because it detects optically thin clouds and is not reliant on albedo or thermal contrast. Thus, our strategy for evaluating polar clouds in CAM relies exclusively on CALIPSO.

3. Results

a. Global evaluation of CAM radiation and clouds

1) CAM BIASES REVEALED BY RADIATIVE FLUX AND CLOUD FORCING COMPARISONS

We begin by contrasting global annual-mean radiative fluxes in CAM4 and CAM5 2001–10 Atmospheric Model Intercomparison Project (AMIP) simulations. The global annual mean top-of-model net radiation balance is positive and similar in both CAM4 ($+1.7 \text{ W m}^{-2}$) and CAM5 ($+1.5 \text{ W m}^{-2}$). That both models achieve a small positive radiative balance for the late twentieth century is not surprising because both CAM versions were developed to produce a stable preindustrial (1850) climate. This similar top-of-model balance is achieved with similar net absorbed solar radiation (CAM4 = 240 W m^{-2} , CAM5 = 239 W m^{-2}), and correspondingly similar outgoing longwave radiation (CAM4 = 238 W m^{-2} , CAM5 = 237 W m^{-2}). Global average fluxes from both CAM versions are within the error bars of satellite-observed top-of-atmosphere (TOA) fluxes, such as those from CERES–EBAF (Loeb et al. 2009).

Clouds have a significant impact on the global radiative balance that is often expressed using TOA shortwave cloud forcing (SWCF) and longwave cloud forcing (LWCF) (Ramanathan et al. 1989). When compared to CERES–EBAF, the global annual mean cloud forcing

biases and rms errors (RMSEs) in CAM4 and CAM5 are remarkably similar. The global annual mean SWCF biases in CAM4 (-1 W m^{-2}) and CAM5 (-2 W m^{-2}) are within 1 W m^{-2} . The global annual mean CAM4 LWCF bias (-1 W m^{-2}) is smaller than the LWCF bias in CAM5 (-4 W m^{-2}). The larger LWCF bias in CAM5 has been partially traced to cold and moist biases in the middle and upper troposphere; however, observational uncertainty has complicated assessment of model LWCF bias magnitudes. Changes in the CERES–EBAF data processing algorithm from version 1.0 to version 2.6 resulted in a 4 W m^{-2} decrease in the global annual mean LWCF from 30 to 26 W m^{-2} , primarily due to a 3 W m^{-2} decrease in clear-sky longwave fluxes from 269 to 266 W m^{-2} . In other words, a CERES–EBAF version change reduced the negative LWCF bias in CAM by 4 W m^{-2} . Because modelers tune to all-sky longwave fluxes and models and observations determine the clear-sky longwave fluxes with significantly different methodologies, LWCF model biases are difficult to diagnose (Sohn et al. 2010).

While climate modelers can adjust the global annual-mean cloud forcing to match observations, geographic cloud forcing patterns are not as easily adjusted. Thus, we next evaluate the geographic distribution of annual mean SWCF and LWCF. Both CERES–EBAF satellite observations and model bias maps are shown in Fig. 1. Like their global annual means, SWCF and LWCF annual mean bias patterns are remarkably similar in CAM4 and CAM5. The shortwave cooling effect of clouds is too strong in the tropics and too weak in the midlatitudes, biases that are common in climate models (Trenberth and Fasullo 2010). The shortwave cooling effect of clouds is also not strong enough in either model in the subtropical transition regimes (e.g., in the transition from stratocumulus to trade cumulus along the GEWEX Cloud System Study (GCSS) Pacific cross section (Teixeira et al. 2011) marked as a gray line on the maps in Fig. 1). The largest obvious bias pattern difference from CAM4 to CAM5 is the removal of a positive LWCF bias in the tropical western Pacific (TWP).

2) EXPOSITION OF COMPENSATING CLOUD FRACTION AND CLOUD PROPERTY BIASES IN CAM

Our experience shows that exposing compensating errors and documenting model cloud biases often requires more sophisticated cloud diagnostics than cloud forcing. For example, given the similarities in their annual mean SWCF and LWCF shown in Fig. 1, it may come as quite a surprise to learn that the global annual-mean gridbox cloud liquid water path (LWP) is almost three-times larger in CAM4 than in CAM5: 127 g m^{-2} versus

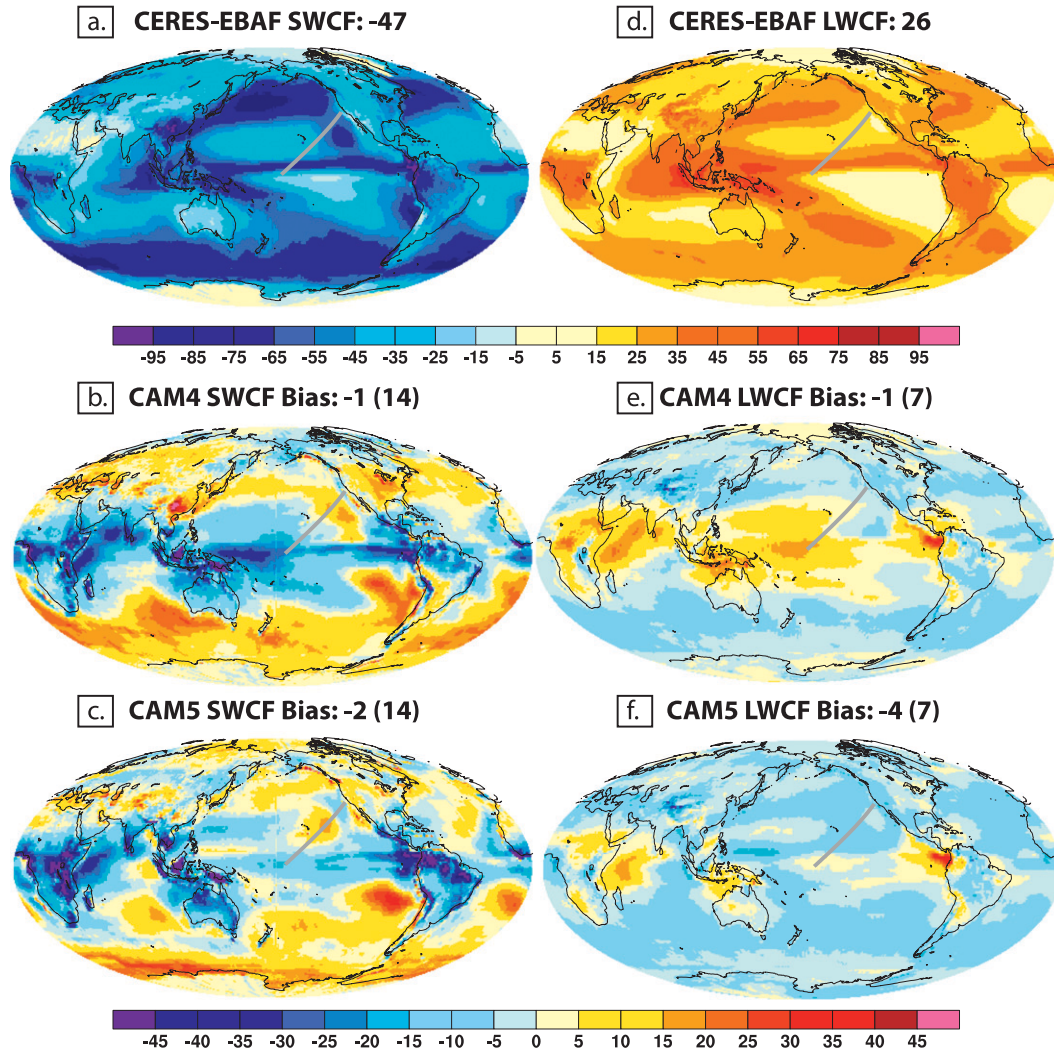


FIG. 1. Global maps of shortwave cloud forcing (SWCF, W m^{-2}) and longwave cloud forcing (LWCF, W m^{-2}): (a) observed CERES-EBAF SWCF, (b) CAM4 SWCF bias (CAM4 – observed), (c) CAM5 SWCF bias (CAM5 – observed), (d) observed CERES – EBAF LWCF, (e) CAM4 LWCF bias (CAM4–observed), and (f) CAM5 LWCF bias (CAM5 – observed). In (a),(d) the value after each panel name reports the global annual mean. In (b),(c) and (e),(f) the value after each panel name reports the global annual mean model bias (CAM – observations) and the value in parentheses reports the RMSE. The solid gray line in each panel shows the GCSS Pacific cross section.

44 g m^{-2} . We use COSP diagnostics to explore these large cloud property differences. We emphasize that all discussed CAM cloud fields are diagnosed using satellite simulators and are compared to their corresponding satellite observations. We do not use *the standard CAM cloud fraction diagnostics* at any point.

We begin by comparing observed and simulator-diagnosed annual-mean total cloud fraction. Figures 2a–c show that the global annual mean total cloud fraction and its spatial distribution are similar in the ISCCP, MISR, and CALIPSO observations. The largest cloud fractions occur in midlatitude storm tracks, tropical convective

regions, and stratocumulus decks, while the smallest cloud fractions occur over subtropical deserts.

Both CAM4 (Figs. 2d–f) and CAM5 (Figs. 2g–i) have annual-mean cloud fraction spatial patterns that broadly match the satellite observations, but it is clear that both CAM versions underestimate global annual mean cloud fraction and have large regional cloud biases. Model cloud deficits are especially pronounced in subtropical stratocumulus and transition regimes, consistent with too weak SWCF in these regions (Figs. 1b,c). Both models have a cloud deficit in the Southern Ocean, consistent with their respective SWCF biases in that region (Figs. 1b,c).

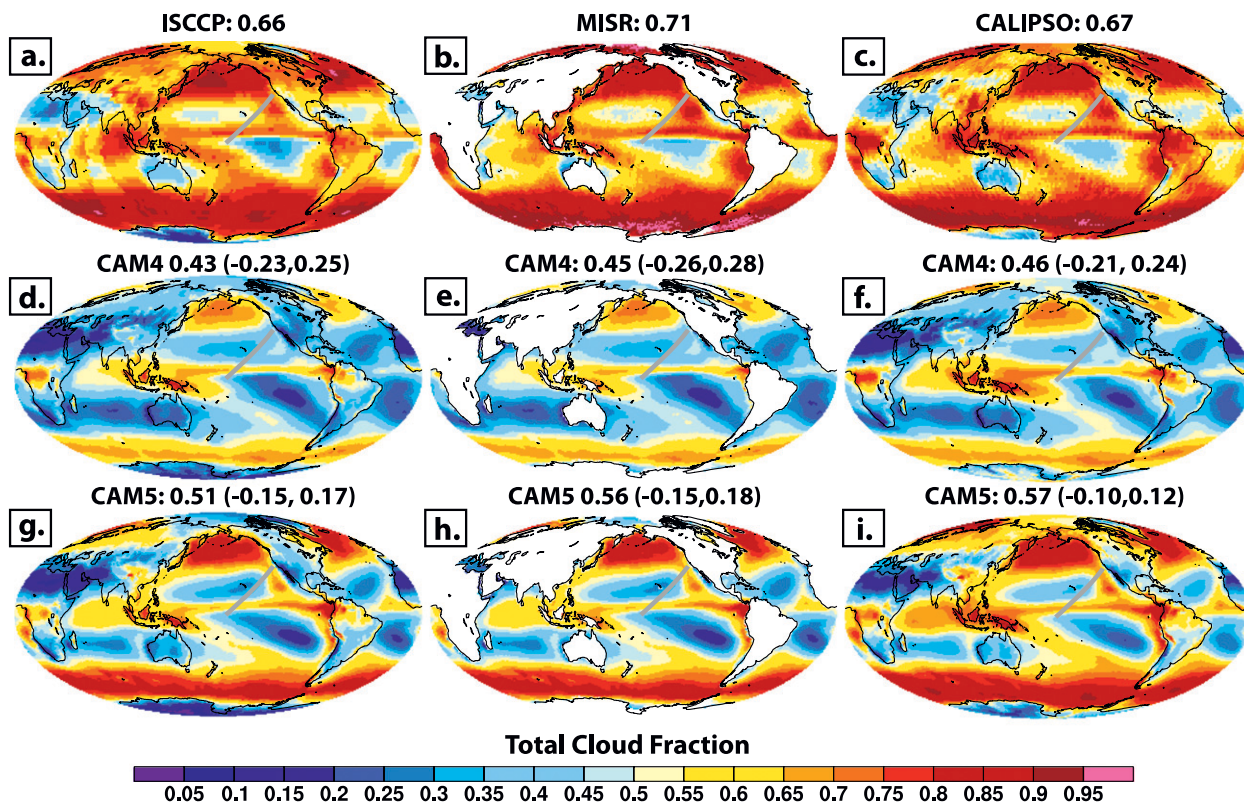


FIG. 2. Global maps of observed and COSP-simulated annual mean total cloud fraction: (a) ISCCP observations ($\tau > 0.3$), (b) MISR observations ($\tau > 0.3$), (c) CALIPSO GOCCP observations ($SR > 5$), (d) CAM4 ISCCP, (e) CAM4 MISR, (f) CAM4 CALIPSO, (g) CAM5 ISCCP, (h) CAM5 MISR, and (i) CAM5 CALIPSO. The value after each panel name reports the global annual mean, while the values in parentheses report the global annual mean model bias (CAM – observations) and model RMSE, respectively. Because they do not include optically thin or broken clouds, the global annual mean cloud fractions observed by MODIS (0.49) and *CloudSat* (0.50) are significantly smaller than those observed by CALIOP, MISR, or ISCCP and are not included in this plot. The solid gray line in each panel shows the GCSS Pacific cross section.

While both models have too little cloud when compared to observations, the biases in CAM5 are significantly smaller than those in CAM4. For all three simulated versus observed cloud fraction comparisons, the global annual mean bias in CAM4 is almost double that of CAM5. In addition, the rms error (RMSE) is substantially larger in CAM4 than in CAM5. Spatially, we see that the larger cloud fractions in CAM5 result from cloud fraction increases in midlatitude storm tracks and subtropical stratocumulus regions, both changes that bring CAM5 cloud fractions closer to the observed cloud fractions.

How do CAM4 and CAM5 produce similar annual mean cloud forcing biases (Figs. 1b–f) with such large differences in their annual mean total cloud fractions (Figs. 2d–i)? The answer lies in the distribution of cloud optical depth (τ). Figure 3a shows observed and simulator-diagnosed global annual mean τ distributions for MISR, MODIS, and ISCCP. This figure shows that CAM4 and CAM5 achieve a similar radiative balance because they have compensating differences in the amounts of

optically thin and optically thick cloud. Consistent with earlier ISCCP simulator-based evaluations, such as the 10 model intercomparison study by Z05, CAM4 has too much optically thick cloud and not enough optically thin cloud, resulting in a flatter τ distribution than is observed. In contrast, CAM5 has a steeper τ distribution that agrees quite well with observations. To the best of our knowledge, CAM5 is the first climate model to match the observed τ distribution with this level of fidelity, which is a significant achievement. The reproduction of observed cloud albedos, which are largely determined by τ , is an important metric of model performance.

Because model resolution can affect cloud properties, it is useful to examine the influence of horizontal resolution on global annual mean τ distributions in CAM4 and CAM5 (Fig. 3b). The ISCCP τ distribution shape differences between CAM4 and CAM5 are similar at both horizontal resolutions examined, but the CAM4 ISCCP τ distribution is more resolution-dependent than that in CAM5. Similar resolution dependence was found

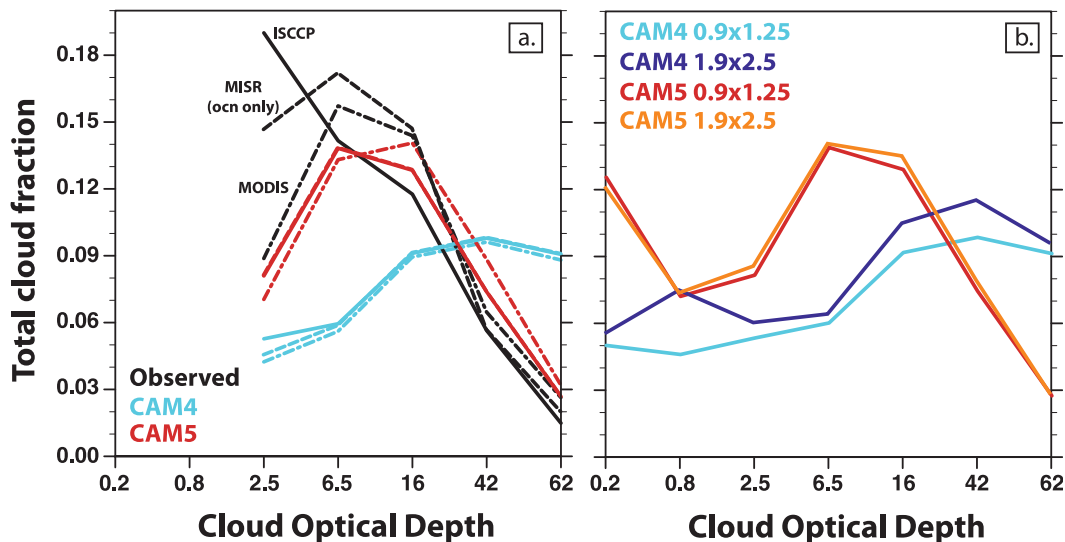


FIG. 3. Global column-integrated cloud optical depth (τ) distributions: (a) MISR, MODIS, and ISCCP from satellite observations, CAM4, and CAM5, (b) ISCCP from CAM4 and CAM5 at both $0.9^\circ \times 1.25^\circ$ and $1.9^\circ \times 2.5^\circ$ horizontal grid resolutions. Below $\tau = 3.6$ there are large intersatellite differences in cloud fraction due to the difference in detection and treatment of cloud edges and subpixel clouds. Observational agreement below $\tau = 1.3$ is especially poor (see P12, Fig. 4) and is therefore not plotted in (a). CAM τ distributions are accumulated over model columns and include the subgrid-scale variability implied by partly cloudy layers.

using the MODIS and MISR τ distributions from CAM4 and CAM5 (not shown).

3) CAM CLOUD BIASES AS A FUNCTION OF SATELLITE, CLOUD OPTICAL DEPTH, AND CLOUD-TOP HEIGHT

We reinforce and expand on the findings shown in Figs. 1–3 by examining global annual cloud fraction bias and RMSE as a function of satellite, τ , and cloud-top height in Fig. 4. Figure 4 shows robust and ubiquitous improvement from CAM4 to CAM5 for all observed versus simulator-diagnosed global annual mean cloud fraction comparisons.

The top row of Fig. 4 compares total cloud fraction for “all cloud,” optically intermediate cloud ($3.6 < \tau < 23$, following the definitions in Z05), and optically thick cloud ($\tau > 23$, again following Z05). As in Fig. 2, total cloud deficits in both versions of CAM are evident, but cloud deficits are smaller in CAM5 than in CAM4. As in Fig. 3, CAM4 has too little optically intermediate cloud and too much optically thick cloud, while CAM5 is a much closer match to the observations.

The vertical distribution of clouds exerts important controls on the longwave energy budget. Thus, we next use Fig. 4 to examine the contribution of high cloud, mid cloud, and low cloud to total cloud amounts and model biases. In both observations and the models, high cloud and low cloud contribute almost equally to the total cloud

fraction, while mid cloud contributes less. Both CAM4 and CAM5 underestimate cloud fraction at all heights, a bias that primarily results from an underestimation of low thin cloud. High, mid, and low cloud fraction biases are similar to their total cloud fraction equivalents and therefore all three contribute to bias reductions from CAM4 to CAM5; however, “all cloud” and optically intermediate cloud bias reductions result mainly from increases in low cloud amount, while optically thick total cloud bias reductions result mainly from decreases in high-topped cloud amount.

Figure 5 provides a spatial context for Fig. 4 by showing maps of CALIPSO cloud fraction bias as a function of height. The figure shows that both models, but especially CAM4, underestimate CALIPSO cloud fraction at all cloud heights in many cloud regimes. Both models share a deficit in low cloud amount on the eastern sides of subtropical oceans at locations where overcast stratocumulus typically transitions to broken trade cumulus clouds. While model cloud deficits are common, there are some regional exceptions. For example, CAM4 has excessive high cloud fractions in the TWP, a regional bias that is improved in CAM5 both in the CALIPSO high cloud fraction bias and in the LWCF bias (Fig. 1).

To identify the regions that contribute to changes in the optically thick cloud from CAM4 to CAM5, Fig. 6 shows maps of MISR optically thick low-topped and MODIS optically thick high-topped cloud observations and model bias. As in Fig. 4, decreases in optically thick

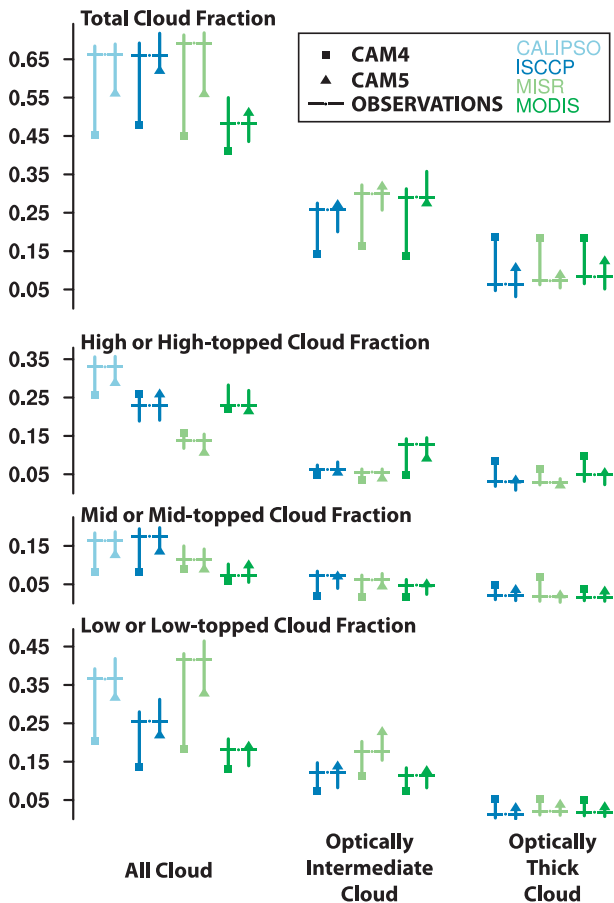


FIG. 4. Observed and COSP-simulated global (60°S – 60°N) annual-mean cloud fractions. Both land and ocean are included except for MISR values, which are ocean only. The satellites and associated simulators, and model versions and observations are coded as shown in the boxed legend. Vertical stems on each model value show the global annual rms error (simulated – observed), directed toward the observed value. Crossbars on each stem occur at the observed mean value, so the global annual bias is the distance from the mean model value (square or triangle) to the crossbar and the distance beyond the crossbar measures the random component of the error. Total cloud fraction shown at the top, followed by high or high-topped (CTH > 7 km or CTP < 440 mb) cloud fraction, mid or mid-topped ($3 \text{ km} < \text{CTH} < 7 \text{ km}$ or $440 \text{ mb} < \text{CTP} < 680 \text{ mb}$) cloud fraction, and low or low-topped (CTH $< 3 \text{ km}$, CTP $> 680 \text{ mb}$) cloud fraction. Unlike the low-topped, mid-topped, and high-topped clouds for MODIS, ISCCP, and MISR, CALIPSO high, mid, and low clouds cannot be summed to obtain the total cloud fraction (see section 2c of the text). Within each panel, the clouds observed by MISR, ISCCP, and MODIS are divided into three column-integrated optical thickness (τ) categories: (left) all cloud, (middle) optically intermediate clouds ($3.6 < \tau < 23$), and (right) optically thick clouds ($\tau > 23$). Differences in detected cloud amount for MODIS, ISCCP, and MISR are discussed in M10 and P12 (see also section 2d of the text). Values from this plot are included in the supplementary material.

cloud fraction bias from CAM4 to CAM5 result more from decreases in high-topped cloud than from decreases in low-topped cloud. Both CAM4 and CAM5 have too much MISR optically thick low-topped cloud in almost all locations, but especially in large-scale descent regions in the tropics and in the midlatitude storm tracks. The CAM4 CALIPSO high cloud bias seen in Fig. 5e is also evident as a MODIS optically thick high-topped cloud fraction bias in Fig. 6e. While MODIS optically thick high-topped cloud bias patterns over tropical land areas share similar patterns in CAM4 and CAM5, positive biases over the Amazon and central Africa are greater in CAM4 than in CAM5.

4) SUMMARIZING CAM CLOUD BIASES IN A SINGLE FIGURE

A Taylor diagram (Taylor 2001) shows correlation, bias, and variability in a single plot, and thus offers a compact visualization of model performance. As such, Taylor diagrams are particularly useful plots to evaluate climate model clouds with the numerous diagnostics available in COSP. We use Taylor diagrams (Fig. 7) to summarize and reinforce the main conclusion of the preceding discussion, namely that the clouds in CAM5 are a closer match to satellite observations than are those in CAM4. Although the LWCF bias is larger for CAM5, the LWCF variability and correlation in CAM5 are closer to observations than they are in CAM4 (Fig. 7a). SWCF statistics are similar in the Taylor diagram for both CAM4 and CAM5, but CAM5 has a better spatial correlation with SWCF observations than does CAM4. While observations constrain total cloud variability to be better in CAM4 than in CAM5, CAM5 has lower bias and better spatial correlation with total cloud observations than CAM4 (Fig. 7b). For CALIPSO clouds, CAM5 has lower bias and better spatial correlation than CAM4 at all heights (Fig. 7c). CALIPSO low and mid clouds in CAM4 have insufficient variance when compared to observations, while those in CAM5 agree much better with observations. As was seen in Figs. 4 and 6, the reduction in optically thick cloud bias from CAM4 to CAM5 is due more to a reduction in high clouds than in low clouds. The variance of optically thick clouds is notably excessive for both models, but more so for CAM4 than for CAM5 (Fig. 7d).

b. The impact of radiatively active snow on CAM5 COSP diagnostics

Radiatively active snow has a substantial impact on simulator-diagnosed cloud amounts in CAM5. It is important to document this impact because it contributes to improved COSP diagnostics in CAM5 relative to CAM4, especially for CALIPSO and *CloudSat* midlevel clouds. The increase in midlevel cloud seen with the

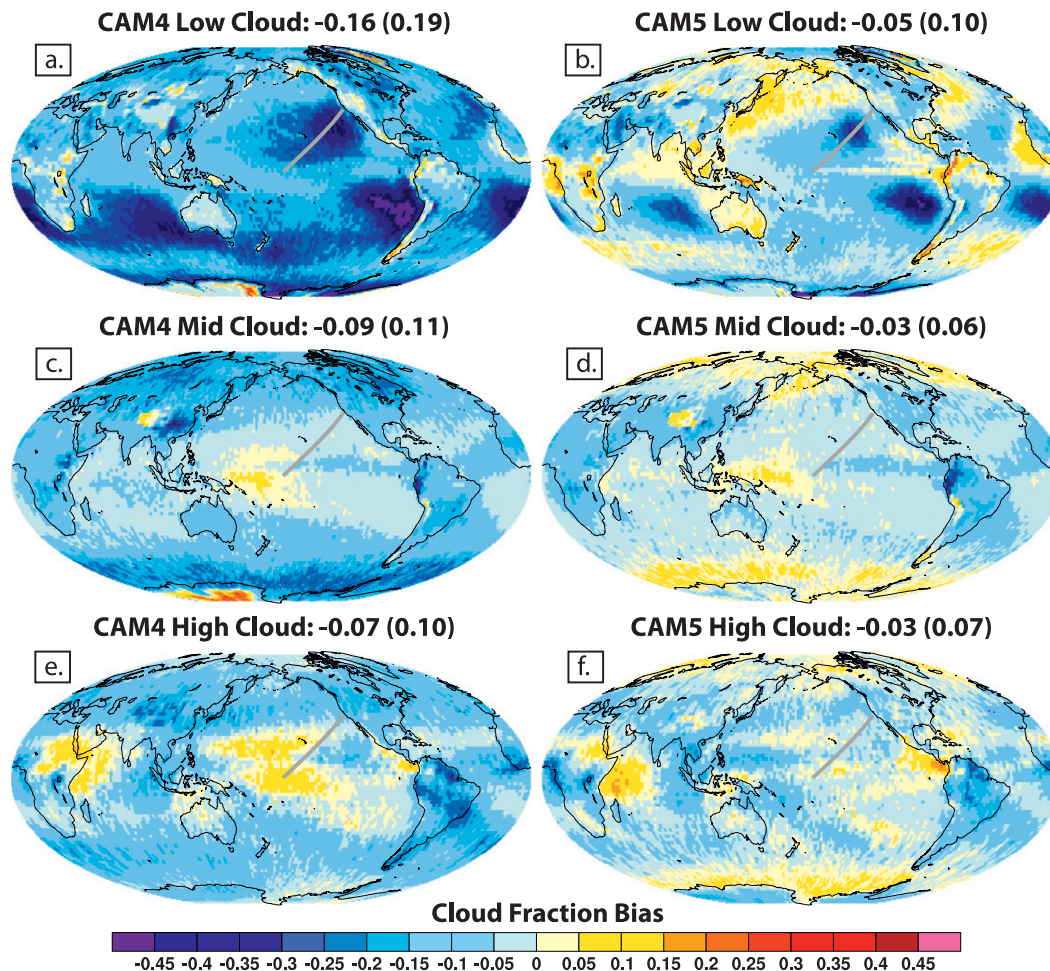


FIG. 5. Global maps of annual mean cloud fraction bias (CAM – CALIPSO GOCCP observations): (a) CAM4 CALIPSO low ($P < 440$ mb) cloud fraction bias, (b) CAM5 CALIPSO low cloud fraction bias, (c) CAM4 CALIPSO mid ($440 < P < 680$ mb) cloud fraction bias, (d) CAM5 CALIPSO mid cloud fraction bias, (e) CAM4 CALIPSO high ($P > 680$ mb) cloud fraction bias, and (f) CAM5 CALIPSO high cloud fraction bias. The value after each panel name reports the global annual mean bias and the value in parentheses reports the RMSE. The solid gray line in each panel shows the GCS Pacific cross section.

passive satellite simulators is smaller than the increases seen with the active satellite simulators (not shown), suggesting that many of the additional midlevel clouds detected by the active satellite simulators are either optically thin or beneath a high-cloud already detected by the passive satellite simulators. Nevertheless, the incorporation of radiatively active snow both into the CAM5 model physics and into COSP diagnostics partially alleviates one of the primary climate model biases documented in Z05 and more recently in B08, namely, the underestimation of model midlevel clouds.

To illustrate the influence of including radiatively active snow in the COSP simulators, Fig. 8 shows zonal annual-mean CALIPSO cloud fraction differences (CAM5 – CAM5 with no snow in COSP). Snow has the largest impact on the CAM5 CALIPSO mid cloud fraction; the

addition of snow results in an increase in global annual mean CALIPSO mid cloud fraction of 0.09, from 0.07 to 0.16, bringing CAM5 closer to the observed value of 0.18. Snow increases the CAM5 CALIPSO high cloud fraction at all latitudes and the CAM5 CALIPSO low cloud fraction primarily at mid and high latitudes preferentially during the winter season. Because snow in CAM5 is only diagnosed when an ice cloud is present at the same or overlying levels, the substantial CALIPSO high cloud fraction increases resulting from snow are surprising.

To further understand the differences in Fig. 8, Figs. 9a and 9b contrast global annual mean CALIPSO scattering ratio (SR) two-dimensional histograms with and without the inclusion of radiatively active snow in the lidar simulator. Consistent with Fig. 8, snow increases lidar-detected cloud ($SR > 5$, Chepfer et al. 2010) the most at midlevels,

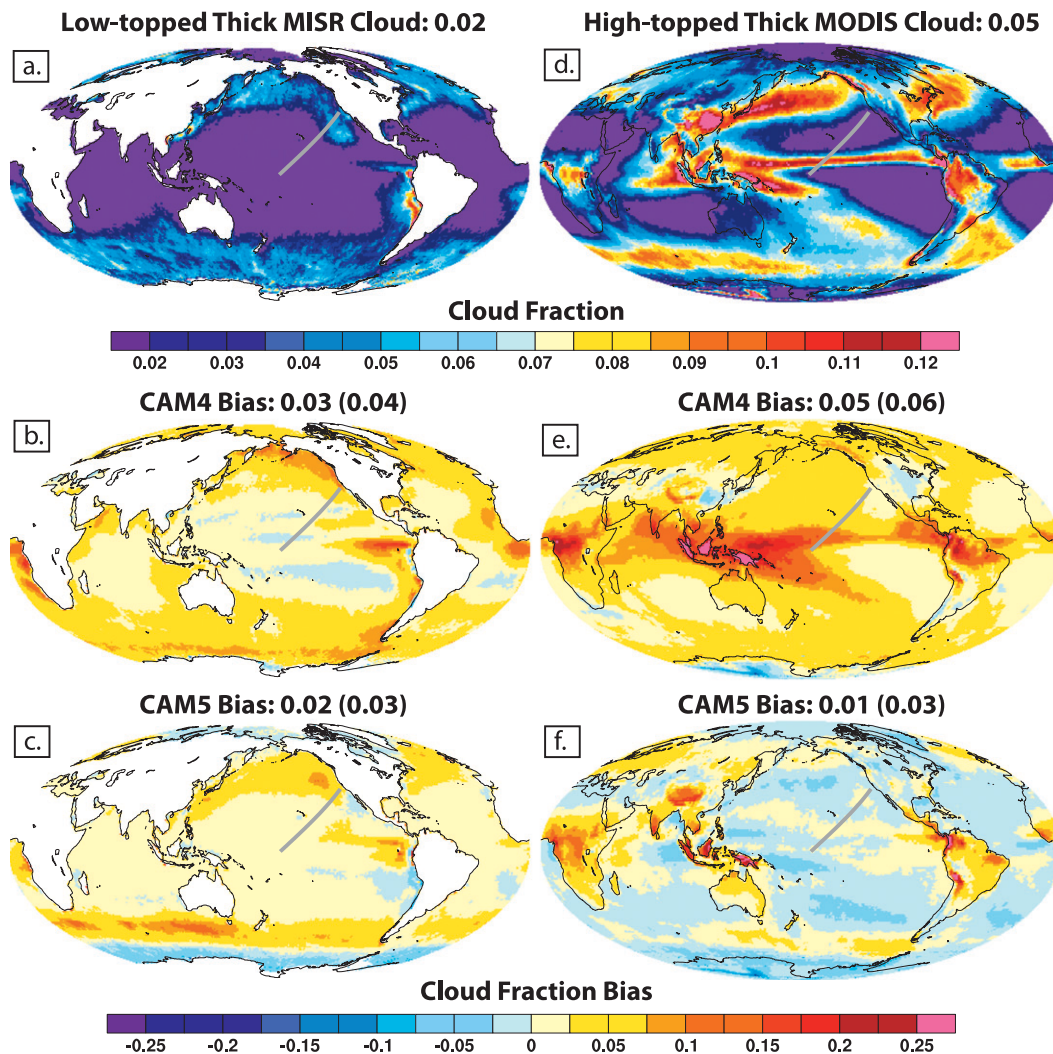


FIG. 6. Global maps of optically thick ($\tau > 23$) annual mean cloud fraction: (a) MISR-observed optically thick low-topped (CTH < 3 km) cloud fraction, (b) CAM4 bias (CAM4 – MISR observed), (c) CAM5 bias (CAM5 – MISR observed), (d) MODIS-observed optically thick high-topped (CTP < 440 mb) cloud fraction, (e) CAM4 bias (CAM4 – MODIS observed), and (f) CAM5 bias (CAM5 – MODIS observed). In (a),(d) the value after each panel name reports the global annual mean. In (b),(c) and (e),(f), the value after each panel name reports the global annual mean bias and the number in parentheses reports the RMSE. The solid gray line in each panel shows the GCSS Pacific Cross Section.

but also at high and low levels. Figure 9 offers an explanation for the surprisingly large increases in CALIPSO high cloud fraction in Fig. 8. Many of the CAM5 high ice clouds have $SR < 5$ and thus do not contribute to the reported CALIPSO high cloud fraction. As a result, the addition of radiatively active snow can have a larger-than-expected influence on CALIPSO high cloud fraction.

c. Regional evaluation of CAM clouds using COSP

1) TROPICAL CLOUD REGIMES

A comprehensive regional evaluation of CAM4 and CAM5 clouds is beyond the scope of this paper, but it is

instructive to highlight select COSP results in a few key regions. We begin by examining clouds along the GCSS Pacific cross section (Teixeira et al. 2011, thin gray line in Figs. 1, 2, 5, and 6), which starts in the heart of the quasi-permanent stratocumulus deck off the California coast, passes through the transition to a cumulus-topped boundary layer, and ends in the intertropical convergence zone (ITCZ). Figure 10 contrasts the vertical distribution of climatological MISR and CALIPSO summer cloud fraction in observations, CAM4, and CAM5 along this cross section, providing a vertical slice through the atmosphere that complements the maps shown in Figs. 1, 2, 5, and 6.

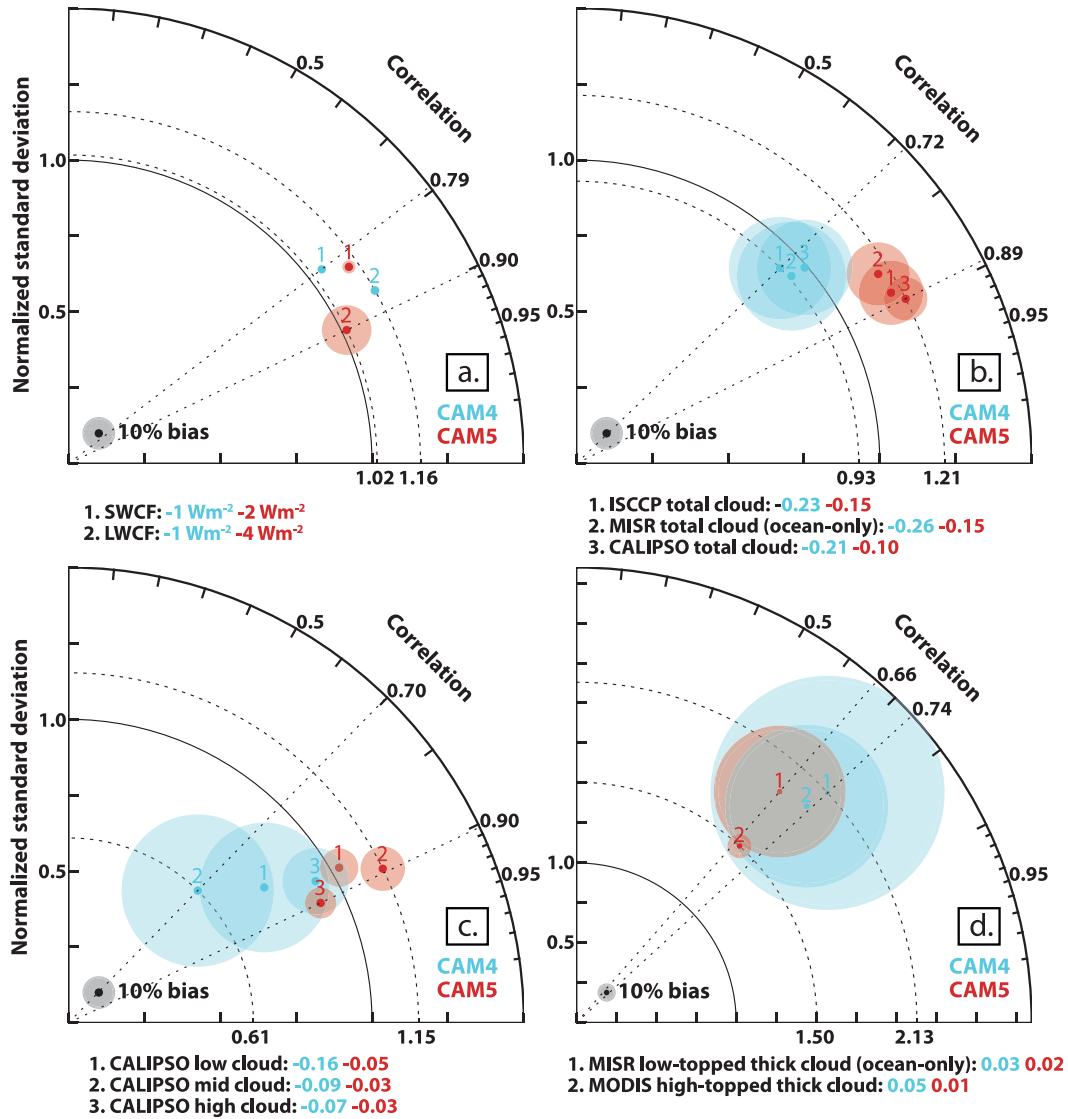


FIG. 7. Taylor diagrams comparing global annual mean CAM4 and CAM5 cloud forcing and cloud properties to observations: (a) SWCF and LWCF compared to CERES–EBAF observations (Fig. 1 shows corresponding maps), (b) ISCCP total cloud, MISR total cloud (ocean only), and CALIPSO total cloud (Fig. 2 shows corresponding maps), (c) CALIPSO low/mid/high cloud (Fig. 5 shows corresponding maps), and (d) MISR low-topped thick ($\tau > 23$) cloud (ocean only) and MODIS high-topped thick ($\tau > 23$) cloud (Fig. 6 shows corresponding maps). The size of the shaded circles is proportional to the global model bias in percent [(model – observed)/observed \times 100]. For reference, the absolute global model bias is included in the legend below each figure panel.

The right side of Fig. 10 begins in the large-scale subsidence regime off the California coast. CAM5 has more stratocumulus than CAM4, which is in better agreement with observations, an improvement also seen in CALIPSO low cloud bias maps (Figs. 5a,b) and in SWCF bias maps (Fig. 1). The increased stratocumulus cloud amounts in CAM5 result from an increase in low-topped $3.6 < \tau < 9.4$ clouds, a change that produces a closer match to the observed τ distribution for CAM5 than for CAM4 (not shown). Despite this improvement

in stratocumulus cloud amount and τ distribution in CAM5 relative to CAM4, the stratocumulus cloud in CAM5 is concentrated too close to the coast when compared to observations. In other words, the transition from stratocumulus to a cumulus-topped boundary layer with smaller cloud fractions occurs too close to the coast in CAM5 when compared to the observations. CAM4 has somewhat different problems with this transition. Unlike the gradual reduction in cloud amount seen in the observations and CAM5, the cloud fraction and vertical

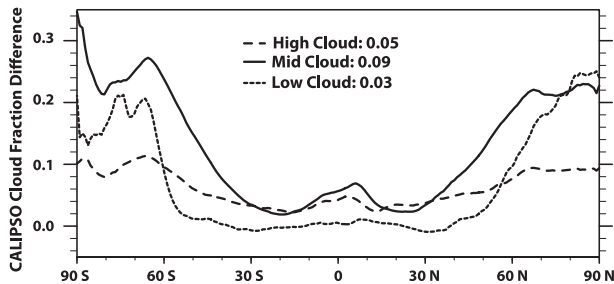


FIG. 8. The influence of radiatively active snow on COSP-simulated cloud fractions shown as zonal annual mean CALIPSO cloud fraction differences (CAM5 – CAM5 with no snow in COSP). The global annual mean cloud fraction differences (CAM5 – CAM5 with no snow in COSP) are reported for CALIPSO high ($P < 440$ mb), mid ($680 < P < 440$ mb), and low ($P > 680$ mb) clouds.

structure from 31.6° to 20° N in CAM4 are uniform. Near 20° N, there is a clear lack of stratocumulus-to-cumulus transition zone clouds in both CAM4 and CAM5, a model bias also evident in CALIPSO low cloud bias maps (Fig. 5), and in SWCF bias maps (Figs. 1b,c). In addition, both models fail to capture the gradual rising cloud top from 36° to 15° N evident in the observations.

In the ascending deep convective regime on the left side of Fig. 10, corresponding to the central Pacific ITCZ, both model versions, but especially CAM4, have too much high cloud. Unlike CAM5, this excessive high

cloud bias is also evident in CAM4 CALIPSO high cloud bias maps (Fig. 5e), CAM4 MODIS optically thick high-topped cloud bias maps (Fig. 6e), and CAM4 LWCF bias maps (Fig. 1e). Comparing CAM4 and CAM5 low cloud biases in this deep convective regime is difficult because the high cloud masks underlying cloud. Finally, although both CAM4 and CAM5 have freezing level cloud (Johnson et al. 1999), the amount is excessive when compared to observations, a bias that may also contribute to excessive CALIPSO mid cloud in the TWP, as seen in Figs. 5c and 5d.

2) ARCTIC CLOUDS

Clouds are critical to Arctic climate feedbacks, but Arctic clouds and their radiative impacts are difficult to observe and evaluate in models (e.g., Kay and Gettelman 2009; Kay et al. 2011). For reasons discussed in section 2d, we rely exclusively on CALIPSO to evaluate Arctic clouds in CAM (Fig. 11). Low clouds are the dominant Arctic cloud type both in the observations and in CAM4 and CAM5. The Arctic cloud fraction seasonal cycle and annual mean cloud fraction are closer to observations in CAM5 than they are in CAM4. Like a recent evaluation of CAM4 Arctic clouds with ground-based and satellite observations that did not use the simulator technique (de Boer et al. 2012), the results in Fig. 11 challenge the observational basis for an ad hoc reduction of low cloud

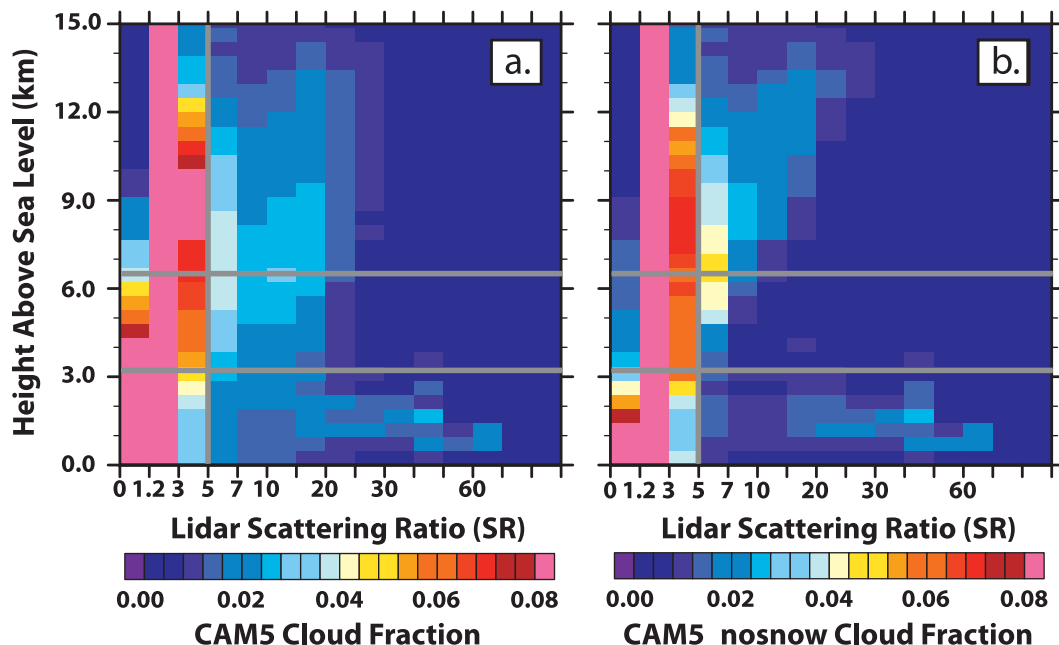


FIG. 9. The influence of radiatively active snow on COSP-simulated cloud fractions: (a) CAM5 global annual mean lidar SR histograms, and (b) as in (a) but with no snow in the lidar simulator. Gray lines on panels show the CALIPSO cloud detection threshold ($SR = 5$), the height boundaries between CALIPSO low and CALIPSO mid clouds (680 mb, ~ 3.2 km), and the height boundary between CALIPSO mid and CALIPSO high clouds (440 mb, ~ 6.5 km).

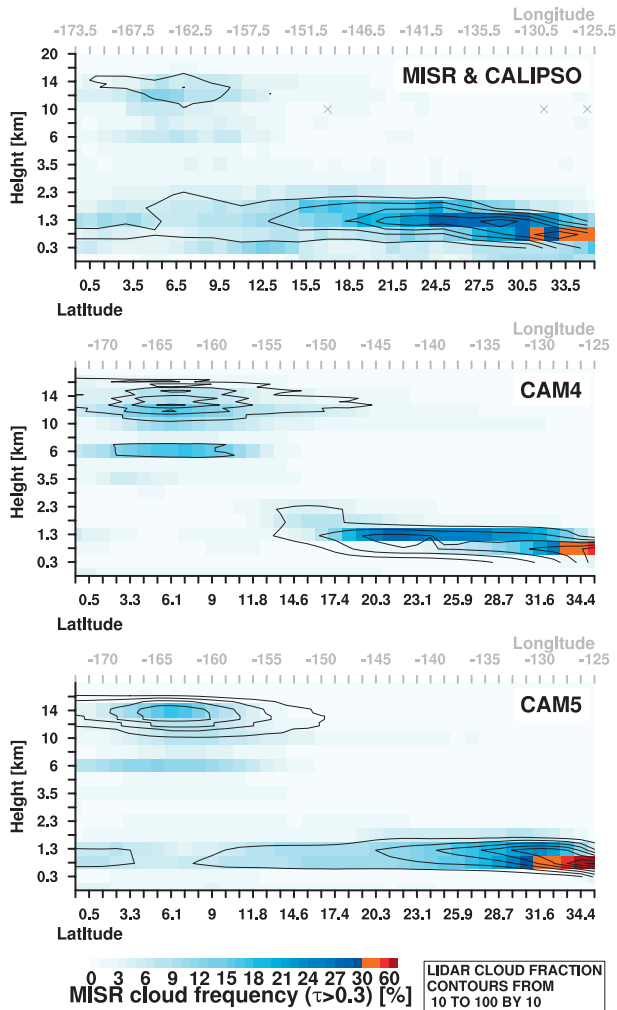


FIG. 10. Climatological June–August cloud fraction along the GCSS Pacific cross section: (top) MISR (color) and CALIPSO (contours) observations, (middle) CAM4 MISR and CALIPSO cloud, and (bottom) CAM5 MISR and CALIPSO cloud. Latitude and longitude points along the cross section are marked at the bottom and top of each panel. The vertical axis shows MISR cloud-top height bins. Three gray markers at 10 km in the top panel mark the locations of the CGILS (CFMIP-GCSS intercomparison of large-eddy and single-column models) transition, stratocumulus, and coastal stratus points.

when there is very little water vapor implemented in CAM4 (Vavrus and Waliser 2008). The small difference between modeled 70°–82°N and 70°–90°N monthly cloud fractions suggest that the lack of CALIPSO observations from 82° to 90°N is not important to these results.

4. Discussion of uncertainty

Our results demonstrate that the synergistic use of cloud observations from multiple satellite platforms and corresponding model-embedded instrument simulators

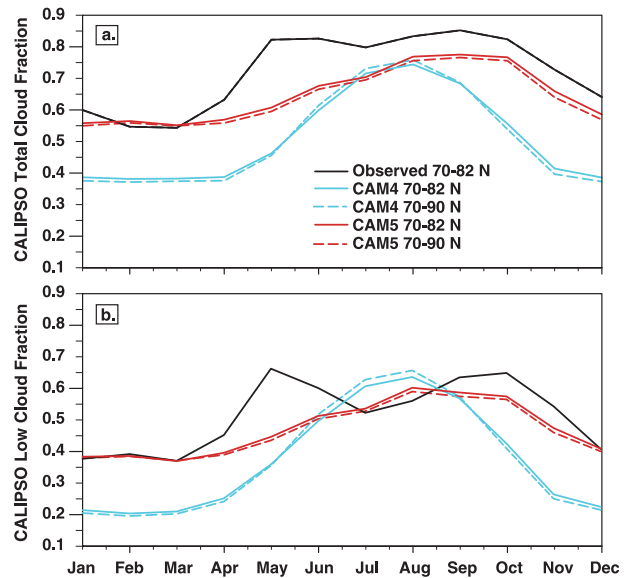


FIG. 11. Monthly evolution of Arctic CALIPSO cloud fraction in observations, CAM4, and CAM5: (a) total cloud fraction, and (b) low cloud fraction.

can robustly identify climate model bias and bias reduction. Yet, both known and unknown uncertainties exist in the application of the instrument simulators for model evaluation. As such, this discussion explains uncertainties that we encountered in the COSP-enabled evaluation of CAM cloud and precipitation fields.

Instrument simulators are an attempt to compute from the model output what a satellite instrument would see if it were “flying” over the model world. Though high cloud masking of low cloud biases and other instrument-specific idiosyncrasies may complicate interpretations, simulator-enabled comparisons are robust when the observational process is well understood and is replicated in the simulator. In contrast, the ability of simulators to reproduce the observational process is limited when the observational process itself is poorly understood. Under these circumstances, simulator-enabled evaluations are, by definition, less robust. Not surprisingly, observational uncertainties frequently plague simulator-enabled comparisons in optically thin and horizontally heterogeneous cloud regimes. For example, there is more scatter in the passive simulator results for $\tau < 1.3$, a known uncertainty whose underlying causes have been explored (P12; M11; M10) and whose importance was taken into account in this study (e.g., Fig. 3). Simulators also have limitations in horizontally heterogeneous cloud regimes. For example, simulators do not take into account the influence of subpixel clouds.

When the reasons for underlying observational differences are not entirely understood, we were cautious

using observations and their corresponding simulators for quantitative model evaluation. Cloud particle sizes derived from MODIS are a model diagnostic currently available in COSP for which observational uncertainty limits the utility of simulator-enabled comparisons. Similar to the Atmospheric Model version 3 (AM3) (Donner et al. 2011), we found a negative liquid drop size model bias in CAM4 and CAM5 when we compared MODIS-simulated and MODIS-observed drop sizes based on retrievals at $2.1 \mu\text{m}$ (Platnick et al. 2003) (not shown). But, the MODIS-observed drop sizes from the retrieval we used are frequently larger than drop sizes from retrieval algorithms based on $3.7 \mu\text{m}$ by other sensors (e.g., Han et al. 1994) or MODIS retrievals at other wavelengths (Zhang and Platnick 2011), especially in conditions where drizzle is present or the retrievals themselves are uncertain (e.g., broken, inhomogeneous, or optically thin clouds). Until the retrieval of liquid drop sizes from MODIS observations is better understood and this information is incorporated into the MODIS simulator, we have justifiable concerns about the ability of the MODIS simulator to provide robust evaluation of CAM particle sizes.

In addition to observational uncertainty, we also found that inconsistent definitions of clouds and precipitation can limit the robustness of simulator-enabled comparisons. For example, our first attempts at COSP-enabled evaluation of CAM5 revealed that radiatively active snow was being included in CAM5, but not in many of the COSP simulators. We remedied this inconsistency by ensuring that all COSP simulators treat snow (Figs. 8 and 9). Our analysis also led us to have concerns about the consistency of the COSP and CAM treatments of precipitation fraction. One simple way to evaluate model precipitation within COSP is to use a radar reflectivity threshold to approximate the occurrence frequency of precipitation [e.g., -15 dBZe , a reflectivity that is an approximate threshold for the presence of precipitation of liquid-phase-only clouds (Liu et al. 2008)]. Similar to results in Stephens et al. (2010) for many climate models and in Zhang et al. (2010) for earlier CAM versions, this crude diagnostic shows that CAM4 and CAM5 overestimate the frequency of precipitation globally and in many precipitation regimes (not shown). Yet, model precipitation fraction is not currently an input into COSP, and COSP assumes that, where precipitation occurs, it occurs over the entire cloud fraction volume. The effect that this assumption has on precipitation-related COSP diagnostics remains unknown, but when invalid this assumption will exaggerate excessive precipitation biases.

Though some uncertainty will always remain, continued research to reduce observational uncertainty, and to quantify the importance of and remove inconsistencies

between model and simulator assumptions, is critical to build trust in simulator-enabled climate model evaluation for future studies. We found a two-way dialog between simulator and climate model developers to be a productive way to reduce uncertainty in simulator-enabled climate model evaluation. Not only did COSP expose CAM cloud biases and highlight inconsistencies in the CAM parameterizations, CAM also exposed COSP deficiencies such as insufficient treatment of radiatively active snow and precipitation fraction.

5. Summary

This study presents a comprehensive global evaluation of climate model cloud biases using COSP instrument simulators and their corresponding satellite observations. The principle finding is that COSP-enabled comparisons robustly show that the physics parameterizations in CAM5 have dramatically reduced three long-standing climate model global cloud biases: 1) the underestimation of total cloud (Fig. 2, Fig. 7b), 2) the overestimation of optically thick cloud (Fig. 3, Fig. 7d), and 3) the underestimation of midlevel cloud (Fig. 7c). The CAM5 midlevel cloud results suggest that climate models underestimate midlevel cloud fraction when the impact of snow on radiative transfer in the atmosphere is neglected. In contrast, CAM4 has large compensating biases in cloud optical properties and cloud amount, biases that are similar to those found in many climate models analyzed using the ISCCP simulator and observations (e.g., Z05). The most striking regional improvements in cloud representation from CAM4 to CAM5 include decreased optically thick high-topped cloud in the deep convective tropical Pacific, increased midlatitude storm track cloud, increased low cloud in the stratocumulus regimes, and an improved seasonal cycle of Arctic clouds. The use of COSP-enabled comparisons was critical for documenting these improvements because, despite having dramatically different cloud fractions and cloud optical properties, compensating errors often allow CAM4 and CAM5 to have similar cloud forcing biases.

Despite the improved representation of clouds in CAM5, cloud biases persist in both CAM versions. Although CAM5 is a closer match to observations, both CAM versions underestimate total cloud when compared to observations. Of particular importance for climate simulations are the significant biases found in subtropical marine boundary layer clouds (e.g., Fig. 10 and complementary maps in Figs. 1b,c and 5a,b). While CAM5 shows some obvious improvements over CAM4, neither model correctly captures the transition from stratocumulus-topped to cumulus-topped boundary layer.

Reasons for these CAM cloud biases have been partially diagnosed in Medeiros et al. (2012) and Hannay et al. (2009). Given this context, it is interesting that Gettelman et al. (2012) report that cloud feedbacks in tropical trade cumulus regions make important contributions to climate sensitivity differences between CAM4 and CAM5 (see their Fig. 9). Other CMIP5 models should be examined to see if similar COSP-identified low subtropical marine cloud biases exist because these clouds have been repeatedly identified as the largest contributors to intermodel spread in climate sensitivity (e.g., Cess et al. 1990; Bony and Dufresne 2005; Williams and Webb 2009; Medeiros et al. 2008).

More generally, this study demonstrates the great diagnostic power of using multiple satellite datasets and simulators to evaluate cloud biases and compensating errors in climate models used for future climate projection. The satellite simulator strategy consistently defines the cloud quantities to be compared, which removes much of the ambiguity that often plagues climate model cloud evaluation. Simulator-enabled comparisons can reliably identify climate model biases when the observational process is both understood and accurately represented within the simulator code. That said, the satellite simulator strategy works best when the satellite observations are most reliable, namely in optically thick horizontally homogeneous cloud regimes. When clouds are optically thin (e.g., $\tau < 1.3$ for passive instruments) and have significant subpixel heterogeneity, simulator-enabled comparisons are less reliable. Above all, when multiple model–observation comparisons yield consistent results (something that frequently occurred in this study), it greatly increases confidence that the identified biases result from model error and not from observational uncertainty. This confidence means comparison of COSP diagnostics with corresponding satellite observations can better illuminate whether climate models produce realistic radiative fluxes and cloud forcing with realistic clouds.

While this study presents a framework for global cloud evaluation in climate models using the numerous new diagnostics in COSP, it leaves a lot to be done. Regional analysis is needed to understand the physical processes responsible for creating the biases identified by this study. Research also is needed to connect COSP diagnostics to cloud feedbacks (e.g., Zelinka et al. 2012). Along these lines, constraining cloud feedbacks may not be possible through evaluation of annual means, which was largely the focus of this study. Ultimately, instrument simulators enable credible intermodel and observed–model comparisons, but using these tools does not reduce our need to understand the essential physical processes, and to incorporate this understanding into parameterizations

that reduce model radiation, cloud, and precipitation biases.

Acknowledgments. JEK, AG, and BE were supported by the U.S. NSF through NCAR. JEK was also partially supported by NASA Grant NNX09AJ05G. BRH and TPA were supported by JPL and the NASA MISR Science Team under Contract NMO710860. SAK, YZ, and JB were supported by Regional and Global Climate and Earth System Modeling Programs of the Office of Science at the U.S. Department of Energy and their contributions to this work were performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. BM was supported by the Office of Science (BER), U.S. Department of Energy, Cooperative Agreement DE-FC02-97ER62402. RP was supported by NASA under Grant NNX11AF09G. We all thank the scientists and software engineers who developed CESM1. Computing resources were provided by NCAR's Computational and Information Systems Laboratory (CISL).

REFERENCES

- Bodas-Salcedo, A., and Coauthors, 2008: Evaluating cloud systems in the Met Office global forecast model using simulated *CloudSat* radar reflectivities. *J. Geophys. Res.*, **113**, D00A13, doi:10.1029/2007JD009620.
- , and Coauthors, 2011: COSP: Satellite simulation software for model assessment. *Bull. Amer. Meteor. Soc.*, **92**, 1023–1043.
- Bony, S., and J.-L. Dufresne, 2005: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.*, **32**, L20806, doi:10.1029/2005GL023851.
- , M. Webb, C. Bretherton, S. A. Klein, P. Siebesma, G. Tselioudis, and M. Zhang, 2011: CFMIP: Towards a better evaluation and understanding of clouds and cloud feedbacks in CMIP5 models. *CLIVAR Exchanges*, Vol. 2, No. 56, International CLIVAR Project Office, Southampton, United Kingdom, 20–24.
- Bretherton, C. S., and S. Park, 2009: A new moist turbulence parameterization in the Community Atmosphere Model. *J. Climate*, **22**, 3422–3448.
- Cess, R. D., and Coauthors, 1990: Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *J. Geophys. Res.*, **95D**, 16 601–16 615.
- Chepfer, H., M. Chiriaco, R. Vautard, and J. Spinhirne, 2007: Evaluation of the ability of MM5 mesoscale model to reproduce optically thin clouds over Europe in fall using ICE/SAT lidar spaceborn observations. *Mon. Wea. Rev.*, **135**, 2737–2753.
- , S. Bony, D. Winker, M. Chiriaco, J.-L. Dufresne, and G. Sèze, 2008: Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model. *Geophys. Res. Lett.*, **35**, L15704, doi:10.1029/2008GL034207.

- , —, —, G. Cesana, J. L. Dufresne, P. Minnis, C. J. Stubenrauch, and S. Zeng, 2010: The GCM-oriented CALIPSO cloud product (CALIPSO-GOCCP). *J. Geophys. Res.*, **115**, D00H16, doi:10.1029/2009JD012251.
- Cole, J., H. W. Barker, N. G. Loeb, and K. von Salzen, 2011: Assessing simulated clouds and radiative fluxes using properties of clouds whose tops are exposed to space. *J. Climate*, **24**, 2715–2727.
- Collins, W. D., 2001: Parameterization of generalized cloud overlap for radiative calculations in general circulation models. *J. Atmos. Sci.*, **58**, 3224–3242.
- de Boer, G., W. Chapman, J. E. Kay, B. Medeiros, M. D. Shupe, S. Vavrus, J. Walsh, 2012: A characterization of the present-day Arctic atmosphere in CCSM4. *J. Climate*, **25**, 2676–2695.
- Donner, L. J., and Coauthors, 2011: The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *J. Climate*, **24**, 3484–3519.
- Gettelman, A., and Coauthors, 2010: Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the Community Atmosphere Model. *J. Geophys. Res.*, **115**, D18216, doi:10.1029/2009JD013797.
- , J. E. Kay, and K. M. Shell, 2012: The evolution of climate sensitivity and climate feedbacks in the Community Atmosphere Model. *J. Climate*, **25**, 1453–1469.
- Hack, J. J., 1994: Parameterization of moist convection in the NCAR Community Climate Model, CCM2. *J. Geophys. Res.*, **99** (D3), 5551–5568.
- Han, Q., W. B. Rossow, and A. A. Lacis, 1994: Near-global survey of effective droplet radii in liquid water clouds using ISCCP data. *J. Climate*, **7**, 465–497.
- Hannay, C., and Coauthors, 2009: Evaluation of forecasted southeast Pacific stratocumulus in the NCAR, GFDL, and ECMWF models. *J. Climate*, **22**, 2871–2889.
- Hogan, R. J., C. Jakob, and A. J. Illingworth, 2001: Comparison of ECMWF winter-season cloud fraction with radar-derived values. *J. Appl. Meteor.*, **40**, 513–525.
- Holtslag, A. A. M., and B. A. Boville, 1993: Local versus nonlocal boundary layer diffusion in a global climate model. *J. Climate*, **6**, 1825–1842.
- Hurrell, J., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary data set for the Community Atmosphere Model. *J. Climate*, **21**, 5145–5153.
- Iacono, M., J. Delamere, E. Mlawer, M. Shephard, S. Clough, and W. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, doi:10.1029/2008JD009944.
- Jakob, C., and G. Tselioudis, 2003: Objective identification of cloud regimes in the tropical western Pacific. *Geophys. Res. Lett.*, **30**, 2082, doi:10.1029/2003GL018367.
- Johnson, R. H., T. M. Rickenbach, S. A. Rutledge, P. E. Ciesielski, and W. H. Schubert, 1999: Trimodal characteristics of tropical convection. *J. Climate*, **12**, 2397–2418.
- Kay, J. E., and A. Gettelman, 2009: Cloud influence on and response to seasonal Arctic sea ice loss. *J. Geophys. Res.*, **114**, D18204, doi:10.1029/2009JD011773.
- , K. Raeder, A. Gettelman, and J. Anderson, 2011: The boundary layer response to recent Arctic sea ice loss and implications for high-latitude climate feedbacks. *J. Climate*, **24**, 428–447.
- Klein, S. A., and C. Jakob, 1999: Validation and sensitivities of frontal clouds simulated by the ECMWF model. *Mon. Wea. Rev.*, **127**, 2514–2531.
- Lawrence, D. M., and Coauthors, 2011: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *J. Adv. Model. Earth Sys.*, **3**, M03001, doi:10.1029/2011MS000045.
- Lin, W. Y., and M. H. Zhang, 2004: Evaluation of clouds and their radiative effects simulated by the NCAR Community Atmospheric Model against satellite observations. *J. Climate*, **17**, 3302–3318.
- Liu, X., and Coauthors, 2011: Toward a minimal representation of aerosol direct and indirect effects: Model description and evaluation. *Geosci. Model Dev. Discuss.*, **4**, 3485–3598.
- Liu, Y., B. Geerts, P. H. Daum, R. McGraw, and M. Miller, 2008: Threshold radar reflectivity for drizzling clouds. *Geophys. Res. Lett.*, **35**, L03807, doi:10.1029/2007GL031201.
- Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the earth's top-of-atmosphere radiation budget. *J. Climate*, **22**, 748–766.
- Mace, G., S. Houser, S. Benson, S. A. Klein, and Q. Min, 2011: Critical evaluation of the ISCCP simulator using ground-based remote sensing data. *J. Climate*, **24**, 1598–1612.
- Marchand, R., T. Ackerman, M. Smyth, and W. B. Rossow, 2010: A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS. *J. Geophys. Res.*, **115**, D16206, doi:10.1029/2009JD013422.
- Medeiros, B., B. Stevens, I. M. Held, M. Zhao, D. L. Williamson, J. G. Olson, and C. S. Bretherton, 2008: Aquaplanets, climate sensitivity, and low clouds. *J. Climate*, **21**, 4974–4991.
- , D. L. Williamson, C. Hannay, and J. G. Olson, 2012: Southeast Pacific stratocumulus in the Community Atmosphere Model. *J. Climate*, in press.
- Morrison, H., and A. Gettelman, 2008: A new two-moment bulk stratiform cloud microphysics scheme in the NCAR Community Atmosphere Model (CAM3). Part I: Description and numerical tests. *J. Climate*, **21**, 3642–3659.
- Neale, R. B., J. H. Richter, and M. Jochum, 2008: The impact of convection on ENSO: From a delayed oscillator to a series of events. *J. Climate*, **21**, 5904–5924.
- , and Coauthors, 2011a: Description of the NCAR Community Atmosphere Model (CAM5). National Center for Atmospheric Research Tech. Rep. NCAR/TN-486+STR, 268 pp.
- , and Coauthors, 2011b: Description of the NCAR Community Atmosphere Model (CAM4). National Center for Atmospheric Research Tech. Rep. NCAR/TN-485+STR, 212 pp.
- Norris, J. R., and C. P. Weaver, 2001: Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3. *J. Climate*, **14**, 2540–2550.
- Park, S., and C. S. Bretherton, 2009: The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model. *J. Climate*, **22**, 3449–3469.
- Pincus, R., S. Platnick, S. A. Ackerman, R. S. Hemler, and R. J. P. Hofmann, 2012: Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. *J. Climate*, **25**, 4699–4720.
- Platnick, S., and Coauthors, 2003: The MODIS cloud products: Algorithms and examples from Terra. *IEEE Trans. Geosci. Remote Sens.*, **41**, 459–473.
- Ramanathan, V., R. D. Cess, E. F. Harrison, P. Minnis, B. R. Barkstrom, E. Ahmad, and D. Hartmann, 1989: Radiative forcing and climate: Results from the Earth Radiation

- Budget Experiment. *Science*, **243**, 57–63, doi:10.1126/science.243.4887.57.
- Rasch, P. J., and J. E. Kristjansson, 1998: A comparison of the CCM3 model climate using diagnosed and predicted condensate parameterizations. *J. Climate*, **11**, 1587–1614.
- Rossow, W. B., and R. A. Schiffer, 1999: Advances in understanding clouds from ISCCP. *Bull. Amer. Meteor. Soc.*, **80**, 2261–2288.
- , G. Tselioudis, A. Polak, and C. Jakob, 2005: Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures. *Geophys. Res. Lett.*, **32**, L21812, doi:10.1029/2005GL024584.
- Schmidt, G. A., and Coauthors, 2006: Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *J. Climate*, **19**, 153–192.
- Sohn, B. J., T. Nakajima, M. Satoh, and H.-S. Jang, 2010: Impact of different definitions of clear-sky flux on the determination of longwave cloud radiative forcing: NICAM simulation results. *Atmos. Chem. Phys.*, **10**, 11 641–11 646, doi:10.5194/acp-10-11641-2010.
- Stephens, G. L., and Coauthors, 2010: Dreary state of precipitation in global models. *J. Geophys. Res.*, **115**, D24211, doi:10.1029/2010JD014532.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106D**, 7183–7192.
- , R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498.
- Teixeira, J., and Coauthors, 2011: Tropical and subtropical cloud transitions in weather and climate prediction models: The GCSS/WGNE Pacific Cross-section Intercomparison (GPCI). *J. Climate*, **24**, 5223–5256.
- Trenberth, K. E., and J. T. Fasullo, 2010: Simulation of present-day and twenty-first-century energy budgets of the Southern Oceans. *J. Climate*, **23**, 440–454.
- Tselioudis, G., Y. Zhang, and W. B. Rossow, 2000: Cloud and radiation variations associated with northern midlatitude low and high sea level pressure regimes. *J. Climate*, **13**, 312–327.
- Vavrus, S., and D. Waliser, 2008: An improved parameterization for simulating Arctic cloud amount in the CCSM3 climate model. *J. Climate*, **21**, 5673–5687.
- Waliser, D. E., J.-L. F. Li, T. S. L'Ecuyer, and W.-T. Chen, 2011: The impact of precipitating ice and snow on the radiation balance in global climate models. *Geophys. Res. Lett.*, **38**, L06802, doi:10.1029/2010GL046478.
- Webb, M., C. Senior, S. Bony, and J.-J. Morcrette, 2001: Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models. *Climate Dyn.*, **17**, 905–922.
- Williams, K., and M. Webb, 2009: A quantitative performance assessment of cloud regimes in climate models. *Climate Dyn.*, **33**, 141–157, doi:10.1007/s00382-008-0443-1.
- Zelinka, M., S. A. Klein, and D. L. Hartmann, 2012: Computing and partitioning cloud feedbacks using cloud property histograms. Part I: Cloud radiative kernels. *J. Climate*, **25**, 3715–3735.
- Zhang, M. H., and Coauthors, 2005: Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *J. Geophys. Res.*, **110**, D15S02, doi:10.1029/2004JD005021.
- Zhang, Y., S. A. Klein, J. Boyle, and G. G. Mace, 2010: Evaluation of tropical cloud and precipitation statistics of Community Atmosphere Model version 3 using *CloudSat* and *CALIPSO* data. *J. Geophys. Res.*, **115**, D12205, doi:10.1029/2009JD012006.
- Zhang, Z., and S. Platnick, 2011: An assessment of differences between cloud effective particle radius retrievals for marine water clouds from three MODIS spectral bands. *J. Geophys. Res.*, **116**, D20215, doi:10.1029/2011JD016216.