Earth and Space Science

RESEARCH ARTICLE
10.1029/2024EA004053

Assessing Clouds in GFDL's AM4.0 With Different Microphysical Parameterizations Using the Satellite Simulator Package COSP

Huan Guo[1], Levi G. Silvers[2], David Paynter[1], Wenhao Dong[1,3], Songmiao Fan[1], Xianwen Jing[4], Ryan Kramer[1], Kristopher Rand[5], Kentaroh Suzuki[6], Yuying Zhang[7], and Ming Zhao[1]

[1]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [2]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA, [3]Cooperative Programs for the Advancement of Earth System Science, University Corporation for Atmospheric Research, Boulder, CO, USA, [4]College of Urban and Environmental Sciences, Hubei Normal University, Hubei, China, [5]SAIC, Science Applications International Corporation, Reston, VA, USA, [6]Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan, [7]Lawrence Livermore National Laboratory, Livermore, CA, USA

Abstract We evaluate cloud simulations using satellite simulators against multiple observational data sets. These simulators have been run within the Geophysical Fluid Dynamics Laboratory's Atmosphere Model version 4.0 (AM4.0), as well as an alternative configuration where a fully two-moment Morrison-Gettelman cloud microphysical parameterization with prognostic precipitation (MG2) is applied, denoted as AM4-MG2. The modeled cloud spatial distributions, vertical profiles, phase partitioning, cloud-to-precipitation transitions, and radiative effects compare reasonably well with satellite observations. Model biases include the under-prediction of total and low-level clouds, especially optically thin/intermediate clouds with cloud optical depth of less than 23, but the over-prediction of thick clouds, indicating "too few, too bright" biases. These biases counteract each other, and give rise to reasonable estimates of cloud radiative effects. The underestimate of low-level clouds is associated with too early and too frequent drizzle/precipitation formation. The precipitation bias is improved in AM4-MG2, where the autoconversion scheme initiates the precipitation more realistically. There also exist discrepancies between models and observations for midlevel and high-level clouds. Additional biases include the underestimate of liquid cloud fraction and the overestimate of ice cloud fraction.

Plain Language Summary We assess the fidelity of cloud simulations using satellite simulators. The simulators facilitate "apples-to-apples" comparisons of the modeled clouds with satellite observations and more clearly identify deficiencies in the modeled clouds within climate models. The simulators are run within the Geophysical Fluid Dynamics Laboratory's Atmosphere Model AM4.0, and its alternative configuration (AM4-MG2). AM4-MG2 is similar to AM4.0 but with a more complicated representation of the physical processes within clouds. Overall, the simulated clouds from both AM4.0 and AM4-MG2 agree well with multiple satellite observations, but this is the result of two counteracting biases. The models produce too few clouds that are overly reflective. AM4-MG2 shows a more realistic initiation of precipitation than AM4.0.

Key Points:

- Cloud simulations from AM4.0 and AM4-MG2 agree well with multiple satellite observations
- Both models over-predict optically thick clouds, but under-predict thin/intermediate clouds
- The precipitation formation is too early and too frequent, which is alleviated in AM4-MG2

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

H. Guo,
huan.guo@noaa.gov

Citation:

Guo, H., Silvers, L. G., Paynter, D., Dong, W., Fan, S., Jing, X., et al. (2025). Assessing clouds in GFDL's AM4.0 with different microphysical parameterizations using the satellite simulator package COSP. *Earth and Space Science*, *12*, e2024EA004053. https://doi.org/10.1029/2024EA004053

Received 20 OCT 2024
Accepted 16 APR 2025

Author Contributions:

Conceptualization: Huan Guo
Data curation: Huan Guo, Wenhao Dong, Ryan Kramer, Kristopher Rand
Formal analysis: Huan Guo, Xianwen Jing
Methodology: Levi G. Silvers, David Paynter, Ryan Kramer, Kentaroh Suzuki, Yuying Zhang, Ming Zhao
Validation: Wenhao Dong, Xianwen Jing
Writing – original draft: Huan Guo
Writing – review & editing: Huan Guo, Levi G. Silvers, Wenhao Dong, Songmiao Fan, Xianwen Jing, Ryan Kramer, Yuying Zhang, Ming Zhao

1. Introduction

Clouds are particularly important for global radiative balance and hydrology. But cloud-related parameterizations remain highly uncertain in general circulation models (GCMs, See Table A1 for acronyms and abbreviations in Appendix A) (IPCC, 2013; IPCC, 2022). Uncertainties in cloud representations as well as how clouds will respond to warming lead to difficulties in estimating climate sensitivity and projecting future climate change (Sherwood et al., 2020). For example, the equilibrium climate sensitivity (ECS), defined as the equilibrium global surface temperature warming resulting from a doubling of atmospheric CO_2 concentration, was estimated to likely lie between 1.5 and 4.5 K a few decades ago (Charney et al., 1979). Currently, based on the most recent CMIP6 (Eyring et al., 2016), ECS is estimated to range from 1.8 to 5.6 K (Zelinka et al., 2020). It has increased from the preceding phase CMIP5 with a range of 2.1–4.7 K, especially on the high end. Moreover this range has been the widest ECS range among the IPCC assessments since the 1990s (Meehl et al., 2020). The discrepancies across GCMs in simulating present-day clouds likely contribute to cloud feedback uncertainties. Therefore, this

growing and large spread warrants a rigorous evaluation of cloud simulations in GCMs against available observations, in order to identify model biases and narrow down their uncertainties.

To evaluate global cloud simulations, satellite observations have proven to be an essential resource, given their near-global coverage and multi-decadal continuous and comprehensive records (Stubenrauch et al., 2013). However, such an evaluation is not straightforward. GCMs simulate temporal and spatial (3D) evolution of cloud micro- and macro-physical properties, along with their background atmospheric state. But satellite instruments directly measure radiance. Based on the radiance measurements, satellite retrievals infer cloud properties from inverse modeling under certain assumptions (Stephens & Kummerow, 2007). As a result, instrument limitations, temporal/spatial sampling uncertainties, retrieval assumptions, different cloud property definitions and scales, among others, challenge a direct comparison between models and observations.

To address this challenge, Cloud Feedback Model Intercomparison Project Observation Simulator Package (Cloud Observation Simulator Package (COSP)) has been developed (Bodas-Salcedo et al., 2011; Swales et al., 2018). COSP comprises a suite of satellite simulators. It is a purely diagnostic software, which is run within a host model to map the modeled clouds to what would be "viewed" by satellite sensors. First, it uses grid box mean profiles of temperature, moisture, hydrometeor mixing ratios, cloud optical properties, as well as surface temperature and emissivity properties, among others from the host model as inputs; Second, COSP splits them into stochastic subcolumns (Klein & Jakob, 1999), and emulates what synthetic observations would obtain; Finally, it generates cloud property diagnostics in a fashion similar to what are used to obtain satellite cloud products. This "model-to-satellite" approach makes it possible that the discrepancies between models and observations reveal actual model biases. COSP has been widely applied in assessing model biases shared by multiple GCMs participating in coordinated experiments (Cesana & Chepfer, 2012; Nam et al., 2012), in identifying biases and proposing improvements across the generations of CAM (Kay et al., 2012, 2016; Medeiros et al., 2023; Zhang et al., 2012), and in evaluating performance and understanding cloud simulation changes in EAM (Zhang et al., 2019, 2024).

However, to our knowledge, few studies have been dedicated to applying COSP for the cloud assessment in GFDL GCMs, for example, the most recent fourth generation atmosphere model (AM4.0). Zhao et al. (2018a, 2018b) systematically analyzed the radiative balance, precipitation, momentum, thermodynamics, effective radiative forcing, climate sensitivity, among others in AM4.0. But they did not provide in-depth analyses on the simulated clouds. This study is motivated to document the performance of AM4.0 in cloud simulations from the COSP perspective. In doing so, we also examine whether the updates in cloud microphysical parameterization will improve the cloud simulations. The cloud microphysical scheme has limited updates from the second generation atmosphere model AM2 to AM4.0. The existing microphysical scheme in AM4.0 has a couple of issues: diagnostic precipitation, lack of explicit treatment of ice number concentration, ice nucleation scheme that often overestimates the number concentration of INP (Fan et al., 2019; Meyers et al., 1992). In order to improve the treatment of precipitation and ice crystal number concentration, as well as aerosol-cloud interactions, we have implemented a fully two-moment Morrison-Gettelman bulk cloud microphysical scheme with prognostic precipitation (MG2) in AM4.0 (Gettelman et al., 2015; Gettelman & Morrison, 2015). Furthermore, to reduce the overestimate of INP number concentration, we have adopted a temperature- and dust-dependent ice nucleation scheme in AM4.0 (Fan et al., 2019). We denote this configuration as AM4-MG2 (Guo et al., 2021). Using the COSP diagnostic outputs, we compare the modeled cloud properties between AM4-MG2 and AM4.0, in order to explore the impacts of cloud microphysical representations. This paper is organized as follows. Section 2 briefly describes the model configurations and satellite observations used in this study. Section 3 evaluates the cloud simulations (including cloud fraction map, vertical profile, phase partition, cloud-to-precipitation transition) from the AM4.0 and AM4-MG2 COSP simulators against observations. The cloud shortwave and longwave radiative effects, and compensating errors are discussed in Section 4. Our findings are summarized in Section 5.

2. Description of Model Configurations and Observations

2.1. AM4.0 and AM4-MG2 Configurations

AM4.0 is GFDL's fourth generation GCM. Its dynamic core adopts the GFDL FV3 hydrostatic approximation (Harris et al., 2020). AM4.0 uses 18 shortwave bands, and shortwave radiation considers CH_4 and N_2O absorption and H_2O continuum, and updates H_2O, CO_2 and O_2 formulations as well (Paynter & Ramaswamy, 2014). For the longwave radiation, the simplified exchange approximation has been applied (Schwarzkopf & Fels, 1991). A "double-plume" convection scheme allows for the coexistence of two plumes. One represents deep convection

Table 1
Differences Between Rotstayn-Klein Cloud Microphysical Scheme and Morrison-Gettelman Microphysical Scheme With Prognostic Precipitation (MG2)

	Moment	Precipitation	Settling hydrometeors	Subgrid distribution
RK	partially two-moment (mass & number of cloud water, only mass of cloud ice)	diagnostic	cloud ice, rain, snow	Triangular, Rotstayn (1997)
MG2	fully two-moment (mass & number of cloud water, ice, rain and snow)	prognostic	cloud water, cloud ice, rain, snow	Gamma, Morrison and Gettelman (2008)

which can penetrate up to the tropopause, while the other represents shallow convection which rarely develops to high altitudes (e.g., above 500 hPa). The entrainment and detrainment rates for these two plumes are formulated in the same way, following Bretherton et al. (2004). However, the fractional lateral mixing rates differ. The lateral mixing rate is a constant in the shallow plume, whereas it is a linear function of column relative humidity in the deep plume. Generally, the shallow plume exhibits much stronger lateral mixing than the deep. This difference in lateral mixing influences the convective depth, precipitation efficiency, and consequently, the amount of condensate in convective precipitation (Zhao et al., 2016, 2018b). The orographic gravity wave drag parameterization has been updated to allow for arbitrary topography (Garner, 2018). The non-orographic gravity wave drag is based on Alexander and Dunkerton (1999). Planetary boundary layers are parameterized following Lock et al. (2000). The Lock scheme here considers down-gradient turbulent diffusion and is applied for convective boundary layers and stratocumulus layers. The large-scale cloud fraction is prognosed according to Tiedtke (1993). The cloud microphysical scheme has been based on Rotstayn (1997). It was originally designed as a one-moment bulk scheme, with prognostic cloud water and ice mass mixing ratios and diagnostic rain and snow. In order to facilitate the study of the aerosol-cloud interactions, a prognostic treatment of cloud droplet number concentration has been introduced (Golaz et al., 2011; Ming et al., 2007). We refer to this microphysical scheme as the Rotstayn-Klein (RK) scheme. Curious readers could find more details about the AM4.0 configuration in Zhao et al. (2018a).

AM4-MG2 is similar to AM4.0, except that the MG2 microphysical scheme replaces the RK scheme. Table 1 summarizes the major differences between these two schemes briefly. MG2 predicts both mass mixing ratios and number concentrations of four hydromemeors: cloud water, cloud ice, rain, and snow. As MG2 prognoses ice crystal number concentration, we have also incorporated an ice nucleation scheme where the number concentration of INP depends on temperature and dust mass concentration (Fan et al., 2017, 2019). Moreover, in order to be consistent with the convective detrainment of ice mass mixing ratio, we have considered the ice number detrainment from convection to large-scale clouds in AM4-MG2, following Kristjansson et al. (2000). Furthermore, both shortwave and longwave radiative effects of precipitating hydrometeors: rain and snow, are considered. More details are available in Guo et al. (2021).

2.2. Satellite Observational Data Sets

To assess the AM4.0 and AM4-MG2 COSP outputs, we have used the climatology of the satellite retrieved cloud properties from MODIS, ISCCP, CALIPSO, MISR, and CloudSat. Table 2 provides the details about the variable,

Table 2
List of Satellite Observations for Evaluating AM4.0 and AM4-MG2 Cloud Observation Simulator Package Simulator Outputs

Observation	Instrument	Variables	Version	Comparison period	Temporal/Spatial resolution
ISCCP	Passive	$\tau - ctp$ Joint histogram (6 τ bins, 7 *ctp* bins)	H-series	200101–201612	monthly, 3-hr/1°
MODIS	Passive	$\tau - ctp$ Joint histogram (7 τ bins, 7 *ctp* bins)	MCD06COSP (V6.2)	200207–202306	monthly/1°
MISR	Passive	$\tau - cth$ Joint histogram (7 τ bins, 15 *cth* bins)	CTH-OD (V7)	200003–202002	monthly/1°
CALIPSO	Active	3D cloud fraction, phase partitioning	GOCCP (V3.1.4)	200606–202205	monthly/2°, L40 (480 m vertically)
CloudSat	Active	radar reflectivity	2B-GEOPROF	200701–201012	monthly/1°
CERES	Passive	SWCRE and LWCRE at TOA	EBAF (Ed4.2)	200003–202302	monthly/1°

temporal/spatial resolution, period, and data version of the above satellite cloud products. Additionally, for the evaluation of CRE (cloud radiative effect), we have adopted the CERES-EBAF Edition 4.2 product.

2.2.1. ISCCP

ISCCP was established in 1982. So far it has delivered the longest, widely used, and well documented record of global distribution and variations of cloud properties (Rossow & Schiffer, 1999). We use the recently developed ISCCP H-series gridded global (HGG) product, which is reported at 1° map resolution and 3 hr time interval. The product is derived from the visible and infrared radiances from imaging radiometers carried on a suite of weather satellite (Rossow et al., 2022; Young et al., 2018). By sampling 3 hr HGG data, we derive the monthly joint histogram of cloud fraction as a function of cloud top pressure (*ctp*) and cloud optical depth (τ), and the climatology of cloud fractions at different *ctp* levels and at different τ bins. Note that from 1983 to late 1990, there is a significant trend of cloud fraction in the ISCCP product. The trend is associated with artifacts, such as satellite zenith angle changes, effective calibration changes, and drifts (Norris & Evan, 2015). In order to minimize its impact, we only analyze the HGG data set during the period of 2001–2016 following Medeiros et al. (2023).

2.2.2. MODIS

MODIS is a 36-channel narrow band imaging instrument flying aboard *Terra* and *Aqua*. Since the launch in the early 2000s, MODIS has produced a broad range of cloud-related observational data sets, for example, cloud fraction, thermodynamic phase, cloud particle size (Pincus et al., 2012; Platnick et al., 2003). We adopt a new global Level-3 data set: MCD06COSP (Pincus et al., 2023). Level-3 data set refers to satellite product that has been processed and mapped on uniform space-time grid scales, usually with self-consistency and completeness. MCD06COSP includes daytime observations from both the *Terra* and *Aqua*. It differs from the regular MODIS products, in that MCD06COSP emphasizes on making more informative comparisons to the MODIS simulator. We use the monthly MCD06COSP data set from 2002 to 2023 with a rectangular latitude–longitude grid of 1° spatial resolution (Table 2). Similar to the ISCCP HGG data set, MCD06COSP provides the joint histogram of cloud top pressure (*ctp*) and cloud optical depth (τ), which is used to obtain the climatology of cloud fraction and its vertical distribution.

2.2.3. MISR

MISR is on board *Terra* that started in 1999 (Diner et al., 2005). Using a stereo-imaging geometric technique, MISR detects cloud top height (*cth*) using with little bias (Marchand et al., 2010). In addition, it has better vertical resolution for liquid-topped midlevel and low-level clouds than ISCCP or MODIS. We apply the MISR version 7 data set. This data set used a newer stereo retrieval algorithm and updated near-infrared calibration to improve the overall data quality (Hillman et al., 2017; Marchand et al., 2010; Mueller & Coauthors, 2017). It contains the joint histogram of *cth* and τ on a global 1° latitude ×1° longitude grid. With the multi-year joint histogram data, we calculate the climatology of total, high-, mid-, and low-level cloud fractions.

2.2.4. CALIPSO

CALIPSO was launched in 2006 and provided data until 2023. It was part of the international "A-Train" constellation and offered global profiling measurements of the atmosphere. CALIPSO was equipped with CALIOP. CALIOP is a two-wavelength (532 and 1,064 nm) polarization lidar instrument. It is able to detect optically thin clouds with optical depth of 0.01 or lower. We employ the CALIPSO GOCCP version 3.1.4 global data set (Chepfer et al., 2010). This data set offers the total, high, middle, and low cloud fractions, cloud vertical profile, and thermodynamic phase partitioning (Cesana & Chepfer, 2013). The version 3.1.4 data set currently covers the period from 2006 to 2022. The horizontal resolution is about 2° latitude x 2° longitude. The vertical height resolution is uniformly 480 m extending from near surface (0.24 km) up to 19 km. Additionally, the CALIPSO GOCCP data set provides cloud fraction as a function of temperature. The temperature bin interval is 3 K.

2.2.5. CloudSat

Similar to CALIPSO, CloudSat was also part of the "A-Train" and equipped with an active sensor. It was launched in 2006 and decommissioned in 2023. It carried a 94 GHz nadir-pointing radar. The radar can penetrate

optically thick hydrometeor layers. This capability offers unique insights on cloud-to-precipitation transition from space. We utilize the radar reflectivity of the CloudSat operational Geometric Profile (2B-GeoProf) product from 2007 to 2011 (Marchand et al., 2008). Using the profile maximum reflectivity, we classify clouds into three categories: non-precipitating, drizzling, and precipitating clouds (Jing et al., 2017; Suzuki et al., 2015), and then estimate the fractional occurrence of precipitation. By combining the radar profile with the MODIS cloud radiative property products, we also analyze the occurrence of radar reflectivity at different in-cloud optical depth and cloud top effective radius (Nakajima et al., 2010; Suzuki et al., 2010).

2.2.6. CERES-EBAF

CERES started in the early 2000s. It has produced long-term and integrated global measurement of the clouds and Earth's radiation budget (Loeb et al., 2018). We utilize the CERES-EBAF Edition 4.2 (Ed4.2) product, which includes many instrument calibration and algorithm improvements. The CERES-EBAF (Ed4.2) product is on a $1° \times 1°$ latitude-longitude grid. The covering period is from 2000 to 2023 here.

These satellite observations have their own strengths and weaknesses. ISCCP, MODIS, and MISR provide passive satellite observations. The ISCCP observation offers the longest observation record since 1983. It has been widely used in climate and weather studies (Klein et al., 2013; Zelinka et al., 2025). MODIS applies more sophisticated retrieval algorithms using more visible and infrared spectral bands. The MODIS measurement is more credible for the high-level clouds, since ISCCP is more likely to mis-identify high topped clouds as the midlevel clouds (Marchand et al., 2010). MISR adopts stereo algorithms that are calibration insensitive for the detection of cloud height. The MISR measurement of the low-level cloud top height is more accurate than ISCCP or MODIS, because temperature inversions could potentially compromise the ISCCP and MODIS retrievals (Kay et al., 2012; Marchand et al., 2010; Pincus et al., 2012).

The active spaceborne CALIPSO lidar and Cloudsat cloud profiling radar bring forth cloud vertical structure products, which are of a special privilege over the passive instruments (e.g., ISCCP, MODIS, and MISR). CALIPSO yields the best satellite measurement for polar clouds, because it is specialized at detecting optically thin clouds and is less dependent on the reflectivity contrast against the underlying surface (Winker et al., 2009). The CloudSat radar reflectivity produces unique precipitating hydrometeor profiles, although they become insensitive to cloud/precipitation particles below 1 km above the surface. Given the limitations from individual instruments and retrieval methods, the different observations could compensate and/or complement each other. Furthermore, multiple observations could be combined to advance our understanding of cloud-related physical processes. For example, Aqua/MODIS and CloudSat are part of the A-Train satellite constellation, which enables simultaneous measurements of cloud and precipitation by different sensors. The combined analyses from two independent measurements can demonstrate how cloud particle size impacts the onset of precipitation (see Section 3.3.2).

3. COSP Simulator Outputs and Evaluations

We examine the outputs from the MODIS, ISCCP, CALIPSO, MISR, and CloudSat simulators that run within the host GCMs: AM4.0 and AM4-MG2. The COSP diagnostics from both GCMs are evaluated against the corresponding satellite observations. For example, we apply the MODIS observation (i.e., MCD06COSP here) to assess the MODIS simulator outputs, and the ISCCP HGG data set to evaluate the ISCCP simulator results.

With AM4.0 and AM4-MG2, we have conducted multi-year AMIP simulations (Gates et al., 1999), where sea surface temperature and sea ice are specified using the HadISST data set (Rayner et al., 2003). The greenhouse gas concentrations, aerosol emissions, and ozone follow the CMIP6 protocols (Eyring et al., 2016). In order to be maximally overlapped with the available satellite observations, our simulation period spans from 1 January 2000 to 31 December 2020. The first-year simulation is ignored. The model climatology for 2001–2020 is analyzed, except that the analysis period for the CloudSat simulator and observation is from 2007 to 2010 (totally 4 years). Both AM4.0 and AM4-MG2 are run at the horizontal resolution of about 1°, and with 33 vertical levels. The model top is around 1 hPa. A sponge layer is applied from the model top to 8 hPa. The time step for model physics is 30 min, while the "acoustic" time step in the dynamical core is 2.5 min.

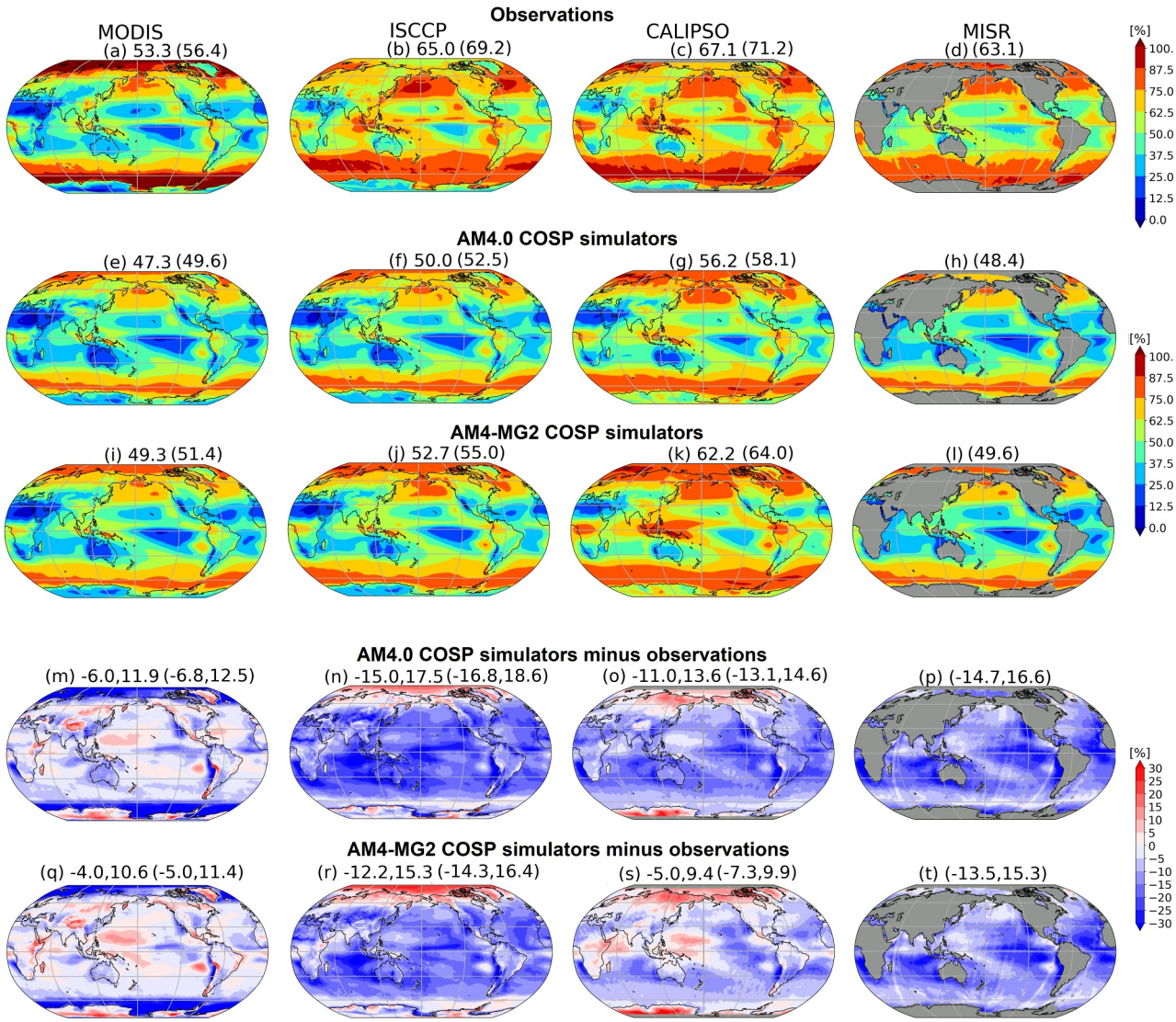

Figure 1. Total cloud fraction from MODIS ($\tau > 0.3$), ISCCP ($\tau > 0.3$), CALIPSO, and MISR ($\tau > 0.3$) observations in (a)–(d), from Cloud Observation Simulator Package (COSP) simulators running within AM4.0 in (e)–(h), from COSP simulators running within AM4-MG2 in (i)–(l), and the differences between AM4.0 COSP simulators and the corresponding observations in (m)–(p), and the differences between AM4-MG2 COSP simulators and the observations in (q)–(t). The top three rows show the annual and global averages from observations and simulators over both land and ocean, and over the ocean only (in parentheses). The bottom two rows present the model biases and root-mean-square errors (RMSEs) over both land and ocean, and over the ocean only (in parentheses).

3.1. Cloud Fraction Maps

3.1.1. Total Cloud Fraction Map

Figures 1a–1d display the climatology of total cloud fraction from the MODIS, ISCCP, CALIPSO, and MISR observations. The total cloud fraction is estimated by summing the joint histogram over cloud optical depth (τ) and cloud top pressure (ctp) or cloud top height (cth) bins for the passive instrument products. Tenuous clouds with $\tau < 0.3$ are neglected because of the large instrument uncertainties (Kay et al., 2012; Marchand et al., 2010). The geographical pattern of the total cloud fraction based on different observations looks similar. There appear more cloudiness over the Southern Ocean (SO), and northern Atlantic and Pacific, but less cloudiness over the Sahara and the tropical regions (excluding the ITCZ). ISCCP, CALIPSO, and MISR show comparable total cloud amounts. Relative to these three, MODIS yields smaller cloud fraction. Over the tropics, their differences approach 10% or more. The primary reason is that the MODIS retrieval excludes partly cloudy pixels (Pincus et al., 2012). The partly cloudy coverage is approximately 15% globally, which largely explains the magnitude

difference of the global averages between MODIS and other three observations. Nevertheless, MODIS reports the largest cloud fraction over the high-latitude ocean (>60°), because MODIS may mis-identify sea ice as clouds especially when its reflectivity is comparable to that of clouds (Chan & Comiso, 2013). In order to reduce MODIS cloud fraction biases, multiple methods have been used, for example, more accurate cloud masking algorithms, multiple spectral bands, and longer wavelengths where surface ice/snow is less reflective (Platnick et al., 2001). Furthermore, combining cloud fraction products from multiple satellite observations with complementary properties has been employed to produce more accurate cloud fraction retrievals spatially and temporarily over the high-latitudes (Liu et al., 2023). The MISR product is only available over ocean surfaces. The MISR cloud fraction is thus absent over the land.

The discrepancies across observations are generally smaller than those between the observations and models. Compared to ISCCP, CALIPSO, and MISR, AM4.0 underestimates the cloud fraction by more than 10% on the global average. AM4-MG2 has some improvements with a smaller cloud fraction bias relative to all four satellite products. Its root-mean-square-error (RMSE) are reduced by about a few percent. Spatially both models suffer from too little cloudiness over the eastern tropical/subtropical Pacific, Atlantic, and Indian Ocean, but too much cloudiness over the polar regions relative to ISCCP and CALIPSO. A persistent model bias lies in the shortage of marine stratocumulus along the west coasts of South America, North America, Australia, and Africa. The deficient stratocumulus problem remains across three generations of GFDL GCMs, as well as many other latest GCMs. With the introduction of MG2 cloud microphysics, such deficiency has been alleviated to some degree, which is consistent with Guo et al. (2021). As shown in Figures 1m–1t, the negative biases are less pronounced (or less "blueish") in AM4-MG2, when evaluated against MODIS, ISCCP, CALIPSO, or MISR.

In addition to the common biases discussed above, the model biases vary when compared to different satellite data sets. For example, both AM4.0 and AM4-MG2 show larger or more negative biases compared to ISCCP or MISR than compared to MODIS or CALIPSO. One reason is associated with optically shallow clouds. If the clouds with $\tau < 1.3$ are excluded, the model biases are reduced by about 5% relative to ISCCP, and by about 10% relative to MISR, respectively. The reduced biases suggest that the shallow clouds are under-predicted in both models, confirming the findings in Zhang et al. (2019). Most shallow clouds are often partially cloudy, which are ignored in the MODIS product (Pincus et al., 2012). Without these shallow clouds, there are smaller model bias changes (by about 2%–3%) with respect to MODIS.

3.1.2. High-, Mid-, and Low-Level Cloud Fraction Maps

We decompose the total cloud fraction to different vertical intervals using *ctp* or *cth*. For MODIS and ISCCP, the decomposition is based on *ctp* using the thresholds of 440 hPa and 680 hPa to separate high-, mid-, and low-level clouds. MISR uses *cth* with the thresholds of 3 and 7 km. CALIPSO adopts the *ctp* thresholds of 440 hPa and 680 hPa (or the *cth* thresholds of 3.2 and 6.5 km). Figure 2 compares the zonally averaged cloud fractions at different vertical levels. The zonal variations of high-, mid-, or low-level fractions from different observations are fairly similar. High-level cloudiness peaks over the tropics and has a second peak over the midlatitudes. Due to the large-scale subsidence (low moisture), high-level cloudiness goes down over the subtropics (near polar regions). For the mid- and low-level clouds, they tend to reach their maxima over the mid-latitudes, and their minima over the subtropics or tropics. While there are broad similarities, remarkable discrepancies across observations exist. The zonal mean high-level cloudiness seems to be less than 20% from the MISR observation, but boost to around 20%–50% based on the ISCCP and CALIPSO products. One reason is that ISCCP includes considerable thin high-level clouds (Marchand et al., 2010). For the midlevel clouds, MODIS and MISR derive much smaller cloud fractions than ISCCP, due to the different retrieval approaches for inferring *ctp* (Marchand et al., 2010; Pincus et al., 2012). Noticeable discrepancies are seen for the low-level clouds, too. MISR observes more partially filled or broken clouds. Consequently the MISR low-level cloud occurs more frequently than the MODIS or ISCCP counterpart by 10% or higher over the subtropics and tropics (Figures 2i, 2j, and 2l). The presence of obvious differences across satellite instruments is not necessarily surprising. With diversity in sensor types, viewing angles and geophysical retrieval methods among others, each instrument is essentially providing a different perspective of the same cloud field. Different satellite products have their strengths and weaknesses, and could be complimentary. This further highlights the importance of using simulators to evaluate a GCM against multiple satellite observations.

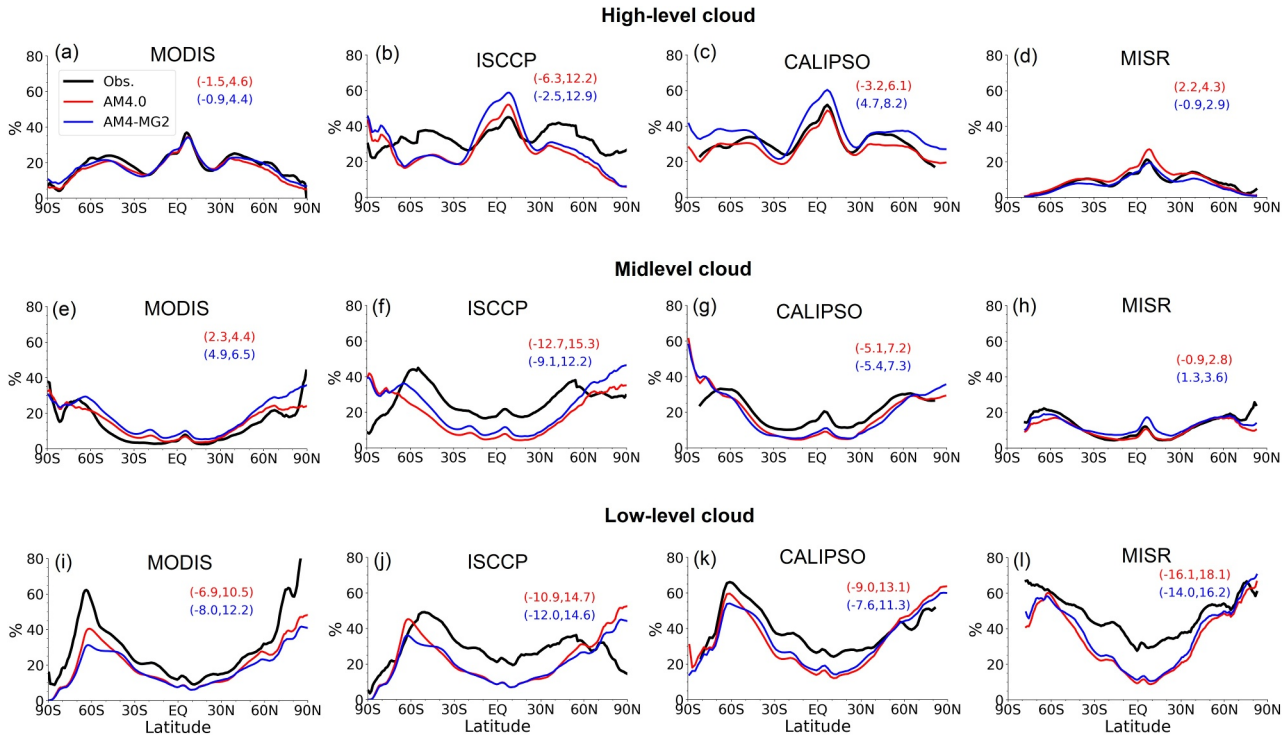

Figure 2. Zonal mean cloud fractions from MODIS ($\tau > 0.3$), ISCCP ($\tau > 0.3$), CALIPSO, and MISR ($\tau > 0.3$) observations (black), from AM4.0 COSP simulators (red), and from AM4-MG2 (blue) simulators for high-level cloud in (a)–(d), for midlevel cloud in (e)–(h), and for low-level cloud in (i)–(l). The AM4.0 (red font) and AM4-MG2 (blue font) model biases and root-mean-squares error (RMSEs) are reported in each panel (in parentheses).

For the low-level clouds, their underestimates in AM4.0 and AM4-MG2 are unambiguously supported by all observations, especially from 60°S to 60°N (Figures 2i–2l), suggesting a "too-few" low-level cloudiness bias as shared by many GCMs (Nam et al., 2012). Poleward of 60°, the satellite retrievals are more uncertain because clouds become more challenging to measure over the highly reflective ice/snow surfaces, especially for the passive instruments. Compared to CALIPSO and MISR, the underestimate of low-level clouds tends to be ameliorated in AM4-MG2.

A more detailed examination of the low-level clouds is conducted by analyzing the cloudiness at different optical depth ranges. Among the three passive instruments (MISR, MODIS, ISCCP), MISR offers more credible retrievals of cloud top height for low-level and midlevel clouds, as well as better identification of trade cumulus (Kay et al., 2012; Marchand et al., 2010; Zhang et al., 2019). So we use the MISR data as observational reference in Figure 3. The MISR low-level clouds are defined as $cth < 3$ km. Obviously optically thin ($0.3 < \tau < 3.6$) and intermediate ($3.6 < \tau < 23$) clouds dominate low-level cloudiness. The thin clouds prevail over the subtropical and tropical cumulus regions. The optically intermediate and thick clouds share similar geographical distributions: more cloudiness over the SO, subtropical stratocumulus region, and Northern Pacific and Atlantic, but less cloudiness over the tropics (Figures 3a–3c). Both AM4.0 and AM4-MG2 seriously underestimate optically thin clouds by about 15%. MISR observes a fair amount of thin clouds (>30%) over the Eastern Atlantic, Pacific, and Indian Ocean, while AM4.0 and AM4-MG2 mostly miss them (<10%). Interestingly, the spatial characteristics of their thin cloud biases are almost the same (Figures 3d and 3g). Both models marginally underestimate the optically intermediate clouds, while moderately overestimate the thick clouds ($\tau > 23$) except for the coastal stratocumulus regions. Comparing their error patterns reveals that AM4-MG2 enhances subtropical marine stratocumulus along the west coasts of continents, and stratus deck over the SO (Figures 3e, 3f, 3h, and 3i). This in turn improves the total cloud fraction bias (Figure 1).

Similarly we evaluate the midlevel cloudiness against MISR observation, and high-level cloudiness against MODIS observation (see Figures S1–S2 in Supporting Information S1), as MISR under-detects high-level cloud amounts (Kay et al., 2012; Marchand et al., 2010; Pincus et al., 2012; Zhang et al., 2019). The error patterns and

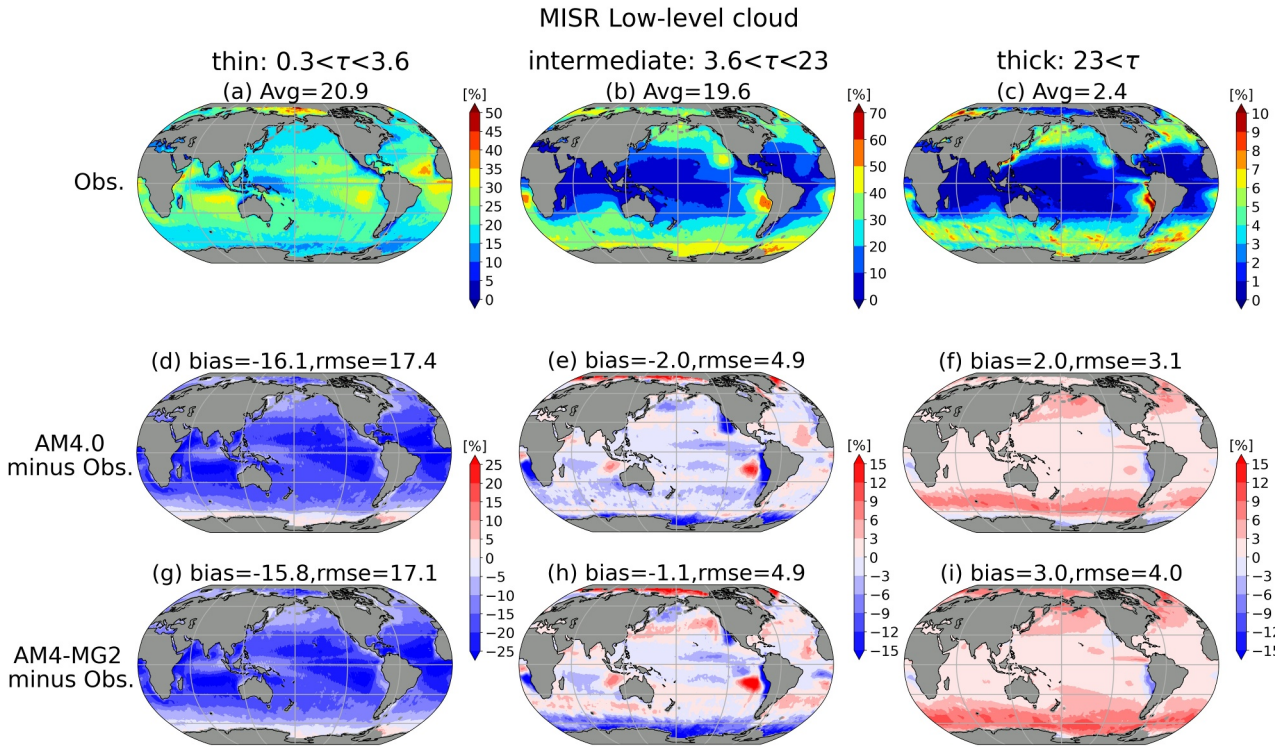

Figure 3. Global maps of optically thin ($0.3 < \tau < 3.6$), intermediate ($3.6 < \tau < 23$), and thick ($\tau > 23$) low-level cloud (cloud top height < 3 km) from MISR observation in (a)–(c), the differences between AM4.0 MISR simulator and observation in (d)–(f), and the differences between AM4-MG2 MISR simulator and observation in (g)–(i). The observational averages (Avg) of the optically thin, intermediate, and thick low-level cloud fractions over the global ocean are presented at the top of panels (a)–(c). The model biases and RMSEs are displayed at the top of panels (d)–(i).

magnitudes between AM4.0 and AM4-MG2 turn out to be close: too few thin/intermediate clouds, but too many thick clouds. These errors persist for high-, mid-, and low-level clouds in both models. We further examine the column-integrated cloud optical thickness distributions. Relative to MISR and MODIS, both models generally under-predicts the optically thin/intermediate cloudiness, and over-predicts the optically thick cloudiness (see Figure S3 in Supporting Information S1). The underestimate of the thin/intermediate clouds also corroborates the bias reductions of total cloud fraction after excluding the thin clouds of $\tau < 1.3$ in Section 3.1.1. The overestimate of the thick clouds implies the over-prediction of highly reflective clouds, or the so-called "too-bright" bias (Nam et al., 2012), which will be discussed further in Section 4.

3.2. Cloud Vertical Profile and Phase Analysis

3.2.1. Cloud Vertical Profile

We examine the annually and zonally averaged vertical profiles of total, liquid, and ice cloud fractions using CALIPSO lidar observation and simulator. Figure 4a shows a variety of climatological regimes of the observed clouds from CALIPSO, which provides detailed cloud vertical information. Over the tropics, as air rises along the ascending branches of the Hadley cell, deep convection develops and clouds can extend from near surface to up to 17 km. The descending branches dissipate clouds in the free troposphere over the subtropics. Over the mid- and high-latitudes, large meridional temperature gradients benefit the development of frontal clouds. Comparing model and observation discloses some common model biases (Figures 4d and 4g). Both models generate excessive high-level clouds relative to the CALIPSO GOCCP product, especially over the tropics and high-latitudes. The excessiveness is amplified in AM4-MG2. But both models are short of mid- and low-level clouds, for example, lack of trade cumulus and coastal stratocumulus as shown in Figure 3. The mixed-phase and liquid clouds over the SO are also significantly underestimated, associated with the lack of the optically thin and intermediate clouds (Figure 3). In addition to the model deficiencies, the shortage of mid- and low-level clouds is partly due to the artifact of the lidar instrument/simulator. The lidar signal usually attenuates quickly,

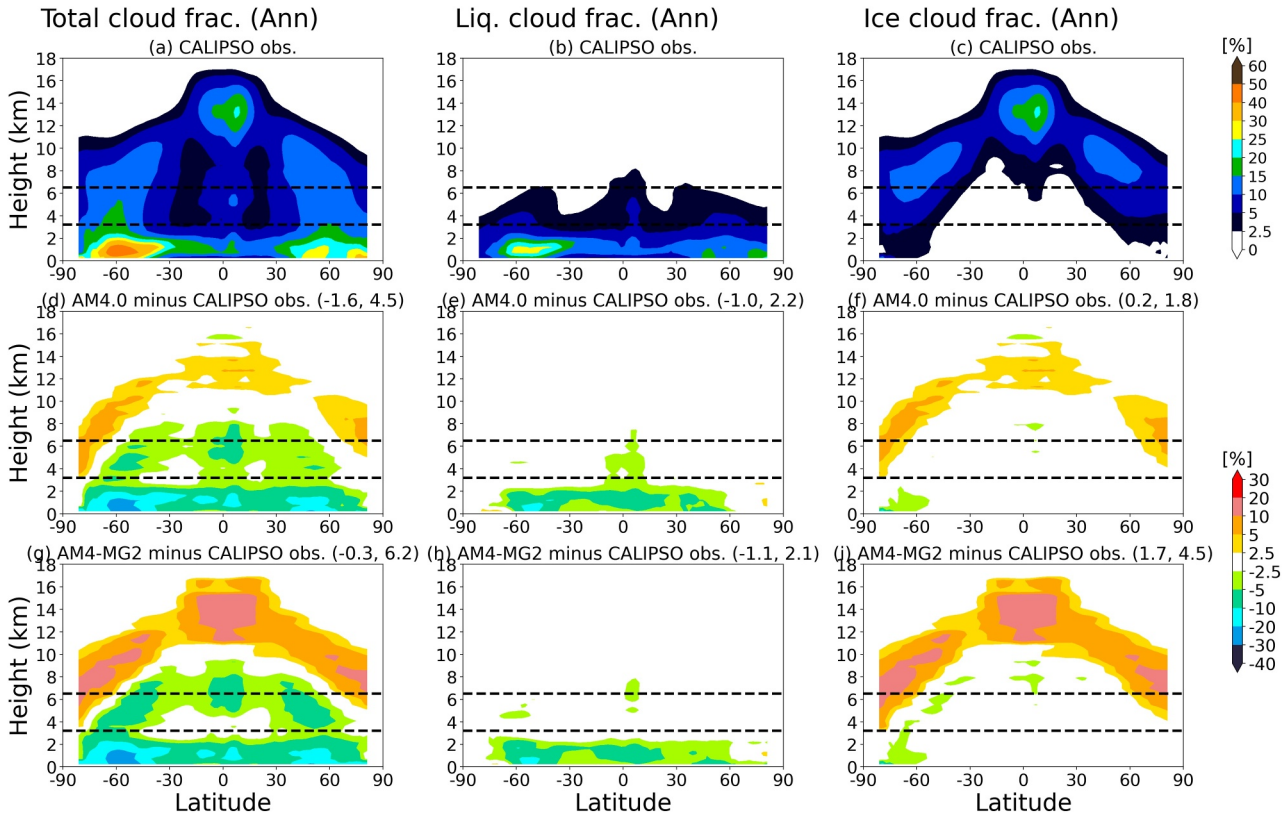

Figure 4. Zonally and annually average total, liquid, and ice cloud fractions from CALIPSO observation in (a)–(c), and the differences between AM4.0 CALIPSO simulator and observation in (d)–(f), and the differences between AM4-MG2 CALIPSO simulator and observation in (g)–(i). The horizontal black dashed lines of 3.2 and 6.5 km separate low-, mid-, and high-level clouds. The model biases and root-mean-squares error (RMSEs) are shown in the bottom two rows (in parentheses).

such that CALIOP is less sensitive to the clouds underneath the top cloud layers. Such behavior is emulated by the lidar simulator, and is likely amplified if the high-level clouds are overestimated. The excessive high-level clouds in AM4.0 or AM4-MG2 diminishes too much lidar signal, and thus potentially masks the underlying clouds and contributes to the underestimate of mid- and/or low-level clouds.

Figures 4b and 4c show the observed liquid and ice cloud vertical profiles, respectively. Apparently liquid clouds are the major component of the low-level clouds (<3.2 km). They are present at all latitudes, and peak over the SO. Ice clouds dominate the high-level clouds (>6.5 km) especially over the tropics, and mostly appear in the free troposphere. Both models overestimate the ice cloud fraction and underestimate the liquid cloud fraction (Figures 4e, 4f, 4h, and 4i). The magnitude of the underestimate of the liquid cloud is smaller than that of the low-level cloud. Similar discrepancy exists for the midlevel cloud (Figures 4d and 4g). The discrepancy is attributable to the "undefined" clouds. The total cloud fraction includes clouds whose thermodynamic phase is undetermined or "undefined", for example, the clouds underneath optically thick clouds (Cesana & Chepfer, 2013). Hence the total cloud fraction may be larger than the sum of liquid and ice cloud fractions. In AM4-MG2, the liquid cloud fraction is underestimated by about 1%, which partially compensates the overestimate in the ice cloud fraction (1.7%). This leads to a reduction in the magnitude of the total cloud fraction bias (−0.3%) but does not necessarily reduce the RMSE. The overestimate of the ice cloud fraction is alleviated in AM4.0. As a result, AM4.0 shows smaller RMSEs in both ice and total cloud fractions than AM4-MG2 (Figures 4d–4i).

Figure 5a shows the total cloud fraction as well as the sum of liquid and ice cloud fraction profiles versus temperature. The difference between the total and the sum of liquid and ice fraction denotes the "undefined" fraction. Clearly there exists non-negligible "undefined" cloud fraction. Its zonal average could be up to 7%, and account for the discrepancies between Figures 4d and 4e, and between Figures 4g and 4h. In both AM4.0 and AM4-MG2, it is not surprising that the mixed-phase (−40°C < T < 0°C) and warm (T > 0°C) clouds are underestimated. The underestimate of the warm clouds is somewhat alleviated in AM4-MG2. This is partially related to

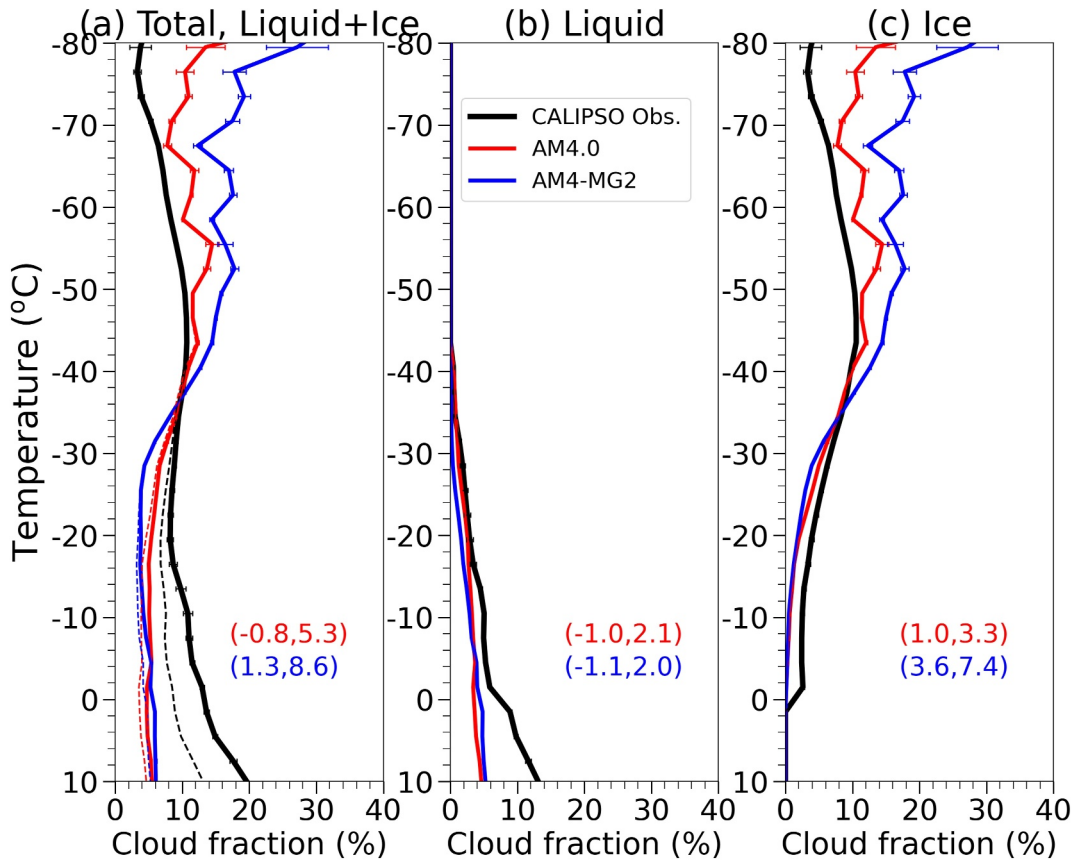

Figure 5. Vertical profiles of annually and horizontally averaged total and sum of liquid and ice (a), liquid (b), and ice (c) cloud fraction with respect to temperature from CALIPSO observation (black), from AM4.0 CALIPSO simulator (red), and from AM4-MG2 CALIPSO simulator (blue). The solid and dashed curves in (a) represent the total cloud fraction, and the sum of liquid and ice (Liquid + Ice) cloud fractions, respectively. The horizontal bars indicate annual standard deviations of total, liquid, and ice cloud fractions. The AM4.0 (red font) and AM4-MG2 (blue font) model biases and root-mean-squares error (RMSEs) are reported in each panel (in parentheses).

the precipitation formation which will be discussed in Section 3.3.1. Both models successfully simulate the monotonic decrease of liquid cloud fraction with temperature, and the occurrence of liquid cloud at temperature as cold as −40°C (Figure 5b). For temperature between −35°C and 0°C, neither model (especially AM4-MG2), has sufficient super-cooled liquid or ice clouds (Figures 5b and 5c), confirming the under-prediction of the mixed-phase clouds (Figure 5a). For temperature colder than −35°C, there seem to be extra ice clouds, particularly in AM4-MG2. The modeled ice cloud fractions are overestimated by about 1.0% and 3.6% for AM4.0 and AM4-MG2, respectively. One possible reason is due to the ice nucleation parameterization. The other possible reason is that the cloudy sample sizes in the models are small at low temperature (Cesana & Chepfer, 2013). The averaged values are sensitive to the fluctuations in cloud fraction occurrences, as supported by relatively large standard deviations in the high-level clouds (Figure 5c).

3.2.2. Cloud Water Phase Partitioning Based on Occurrence Frequency

Figure 6 illustrates the ice cloud fraction percentage as a function of temperature. The ice cloud percentage is defined as the ratio of the ice cloud fraction over the total cloud fraction. Note that the ice percentage here is the cloud occurrence frequency ratio, instead of the mass ratio of cloud condensate. Overall, the modeled ice percentages compare favorably with the observation on the global scale (solid lines). The temperature ranges for the co-existence of liquid and ice phases in AM4.0 and AM4-MG2 span from −40°C to 0°C, which overlaps with the CALIPSO GOCCP observation. The temperatures for the ice percentage of 50% are −21.5°C and −20.3°C for AM4.0 and AM4-MG2, respectively. These match well with the observed value of −20.6°C. When temperature is warmer than −20°C, both models tend to underestimate the ice percentage, largely because the ice cloud fractions

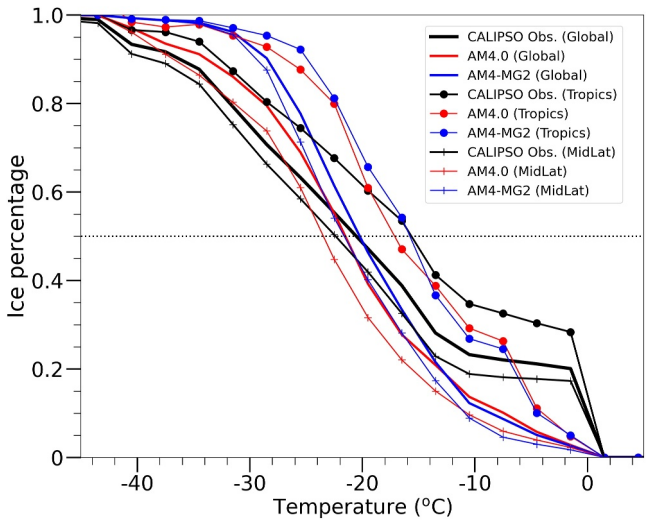

Figure 6. Ice cloud fraction percentage versus temperature diagnosed from CALIPSO observation (black), from AM4.0 CALIPSO simulator (red), and from AM4-MG2 CALIPSO simulator (blue) over global domain (no marker), over tropics [−30S, 30N] (dot maker), and over mid-latitudes [60S, 30S] and [30N, 60N] (plus marker). The dotted black horizontal line indicates the 50% ice percentage.

are biased low (Figure 5c). With a quick increase in ice fraction at temperature colder than −20°C, both models tend to over-predict the ice percentage. We also calculate the ice percentages over the tropics (solid lines with dot markers) and midlatitudes (solid lines with plus markers). The geographical variations do not necessarily alter the relationship between ice percentage and temperature. The ice percentage is generally higher over the tropics than over the globe, but it turns slightly lower over the mid-latitudes.

3.3. Cloud-To-Precipitation Transition

3.3.1. Precipitation Occurrence

As shown in Figures 1 and 2, both AM4.0 and AM4-MG2 underestimate the cloud fraction, especially low-level cloud fraction. Because precipitation is a dominate sink removing cloud water, we evaluate the fractional occurrence of precipitation, which quantifies how likely clouds are to develop precipitation. The CloudSat observation and simulator provide the radar reflectivity, which enables us to differentiate precipitating particles from cloud particles. Following previous studies (Haynes & Stephens, 2007; Kubar et al., 2009; Suzuki et al., 2015), we classify clouds into three categories using the maximum radar reflectivity (Z_{max}):

1. "Non-precipitating" if -30 dBZ $< Z_{max} < -15$ dBZ,
2. "Drizzling" if -15 dBZ $< Z_{max} < 0$ dBZ,
3. "Precipitating" if $Z_{max} > 0$ *dBZ*.

The fractional occurrence (f_i) for each category is defined as $f_i = \frac{n_i}{n_{tot}}$, where i refers to non-precipitating, drizzling, or precipitating clouds, and n_i is the sample number of each category, and n_{tot} is the total sample number of clouds in a pixel (or gridbox) from the CloudSat radar observation (or simulator). Cloudsat radar has a minimum detectable signal of around −30 dBZ. The thresholds of −15 dBZ and 0 dBZ are determined from observations and models, which correspond to precipitation rates of about 0.25–0.5 and 2–5 mm/day, respectively (Wood et al., 2009).

Figure 7 illustrates the spatial distribution of the fractional occurrence of precipitation between 60°S and 60°N averaged from January 2007 to December 2010. Because there was only day-time CloudSat observations after 2011, our analysis was limited to prior to that time. Although it is shorter than other satellite analysis periods in this study, we do not expect the climatology of the precipitation occurrence will change qualitatively. As shown in the CloudSat observations (Figures 7a–7c), non-precipitating clouds are prevalent. Over the mid-latitudes, up to 80%–90% of clouds do not precipitate. From the subtropical land to the tropical ocean, it becomes more likely to drizzle or rain. More drizzling clouds are present over the subtropical ocean, especially over the shallow cumulus region. Toward the ITCZ and Indian ocean, stronger precipitation develops, and precipitating clouds gradually take over. On average, the non-precipitating clouds occur most frequently (52%), followed by the drizzling clouds (32%). The precipitating clouds occur least frequently (16%).

Nevertheless, the radar simulators running within AM4.0 and AM4-MG2 yield different results. Both models generate precipitation too often. As shown in Figures 7d–7i, the drizzling and precipitating occurrences are over-predicted, while non-precipitating clouds are under-predicted. In AM4.0, drizzling clouds prevail. More than 50% of clouds produce drizzle over the mid-latitude ocean. Over the tropics, the drizzling occurrence is about 40%–50%. Over the ITCZ and Indian ocean, the precipitating occurrence approaches 40%–50% or even higher. It is notable that the precipitation occurrence climbs to 80%–90% over the east tropical Pacific, which effectively dissipates clouds and leads to the low biases in cloudiness (Figures 1m–1t).

Compared to AM4.0, AM4-MG2 improves the precipitation occurrence (Figures 7g–7i). The chance of non-precipitating clouds increases by about 20%, while drizzling and precipitating clouds decrease by about 10% each. Over the mid-latitudes and sub-tropics where the large-scale stratiform precipitation dominates, the spatial features of drizzling and precipitating occurrences in AM4-MG2 resemble AM4.0, except for reduced magnitudes. Over the tropics, AM4-MG2 also exhibits less drizzling and precipitating frequencies than AM4.0. This

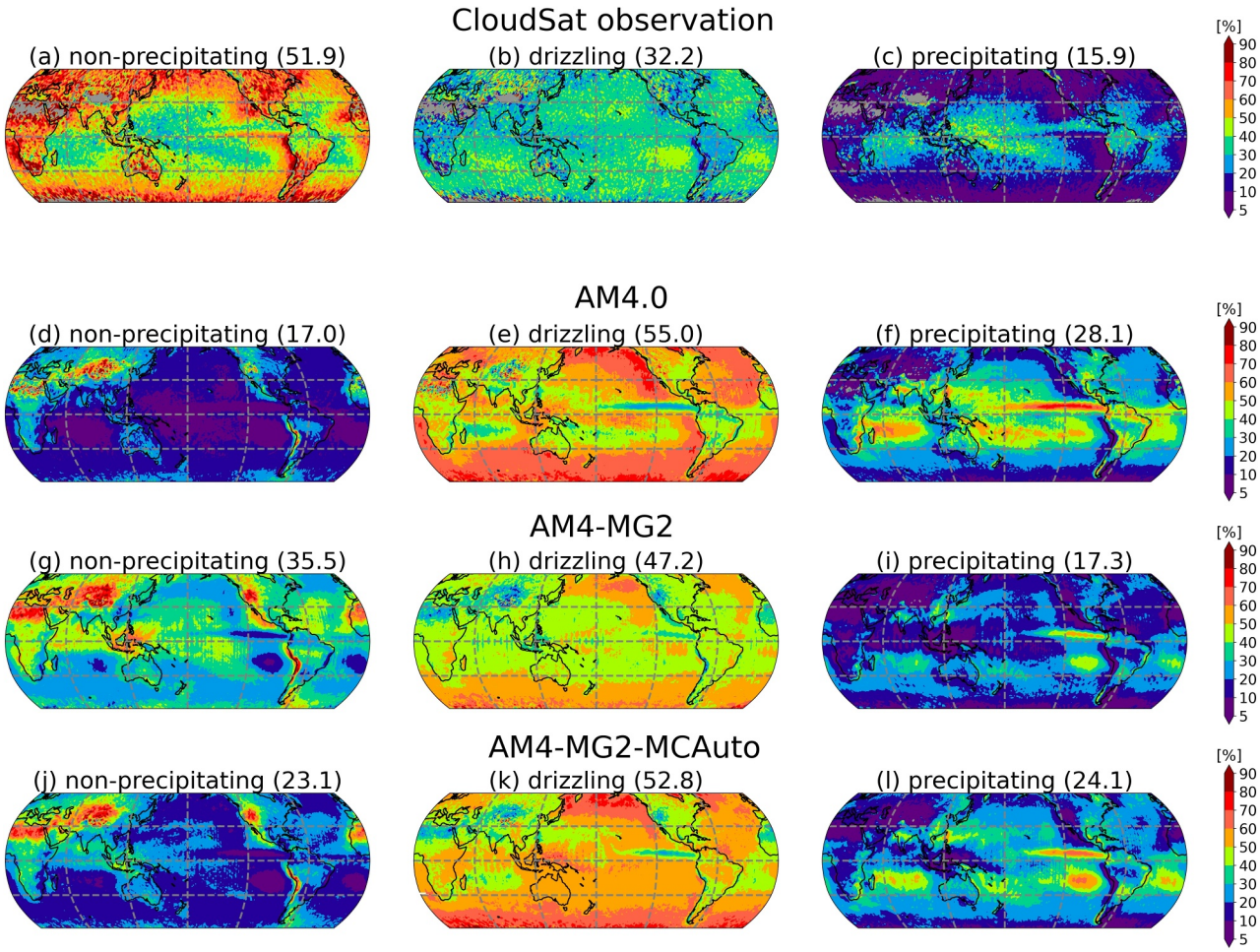

Figure 7. Maps of annually averaged non-precipitating (maximum column radar reflectivity: $Z_{max} < -15$ dBZ), drizzling (-15 dBZ $< Z_{max} < 0$ dBZ), and precipitating ($Z_{max} > 0$ dBZ) cloud occurrence frequencies over [60°S, 60°N] from CloudSat observation in (a)–(c), from AM4.0 CloudSat simulator in (d)–(f), from AM4-MG2 CloudSat simulator in (g)–(i), and from AM4-MG2-MCAuto CloudSat simulator in (j)–(l). Note that AM4-MG2-MCAuto is the same as AM4-MG2 except that the Manton and Cotton (1977) autoconversion scheme is adopted, instead of the Seifert and Beheng (2001) scheme. The spatial average from 60°S to 60°N is presented in the parenthesis.

suggests less efficient precipitation formation. As confirmed by Guo et al. (2021), the precipitation efficiency is reduced from 0.66 in AM4.0 to 0.56 in AM4-MG2. We have increased the lateral mixing rate of the deep plume in the "double-plume" convection parameterization from 0.9 in AM4.0 to 1.1 in AM4-MG2 to enhance stratiform precipitation. The increase turns out to have little impact on the frequency of drizzle and precipitation formation.

As autoconversion is responsible for initiating precipitation, we conduct a sensitivity test where the Manton and Cotton (1977) autoconversion parameterization, which is used in AM4.0, is applied within the AM4-MG2 configuration. We denote this test as AM4-MG2-MCAuto. Note that the default autoconversion scheme in AM4-MG2 follows Seifert and Beheng (2001). It is clear that the Manton and Cotton (1977) scheme boosts the precipitation development (Figures 7j–7l). The average drizzling and precipitating occurrences rise by 5% and 7%, respectively, as the non-precipitating frequency drops by 12%. More than that, the Manton and Cotton (1977) remarkably enhances the similarity of the geographical characteristics of the drizzling and precipitating occurrences between AM4.0 and AM4-MG2-MCAuto. The similarity highlights the critical role of the autoconversion representation for capturing the cloud-to-precipitation transition. Note that the representations of other microphysical processes (e.g., rain evaporation, accretion) also play a role for the modeled precipitation occurrences.

3.3.2. Contoured Frequency by Optical Depth Diagram (CFODD)

Another metric for examining the cloud-to-precipitation transition is Contoured Frequency by Optical Depth Diagram (CFODD) (Nakajima et al., 2010; Suzuki et al., 2010). This metric assesses the vertical structure of the radar reflectivity with respect to optical depth at different hydrometeor particle sizes. Here we analyze the CFODD over the global ocean. The observed cloud top effective radius R_e and cloud optical depth are obtained from Aqua/MODIS shortwave measurement, and the observed reflectivity is provided by the CloudSat cloud profiling radar. Both Aqua and CloudSat flew in the A-Train constellation, and their products matched closely in time and space (Stephens et al., 2002). These two independent pieces of measurement information are combined to generate the statistics shown in Figures 8a–8c. With increasing cloud top effective radius R_e, the observation-based statistics illustrate progressive transitions from non-precipitating (Figure 8a) through drizzling (Figure 8b) to precipitating (Figure 8c) clouds. Unfortunately, neither AM4.0 nor AM4-MG2 can capture this monotonic transition. Regardless of R_e, there always exists a higher chance of precipitating occurrence (reflectivity > 0 dBZ) than observed. Even at the smallest R_e range (5–10 μm), the reflectivity peaks around 5 dBZ or higher (Figures 8d and 8g). However, the satellite observations indicate that the precipitation formation is not efficient until R_e reaches 15 μm or larger (Figure 8c). This discrepancy with observations suggests that the onset of precipitation in both models begins too early in the process of hydrometeor growth. The mean radar reflectivity generally increases with cloud optical depth, and the increase is more significant with larger R_e (see black curves in Figures 8a–8c). But the modeled increase of the mean radar reflectivity shows less sensitivity to R_e (black curves in Figures 8d–8i).

Moreover, both models exhibit higher probability for drizzling or precipitating clouds (reflectivity > −15 dBZ) than the observations, implying that the occurrences of precipitation are too frequent. Compared to AM4-MG2, AM4.0 is more likely to precipitate. One reason is due to the Manton and Cotton (1977) autoconversion scheme. Its application in the sensitivity test of AM4-MG2-MCAuto leads to stronger peaks (or higher chances) in drizzling and precipitating categories for all three R_e ranges (Figures 8j–8l). Guo et al. (2021) assessed the instantaneous autoconversion process rates from AM4.0 and AM4-MG2 against the in-situ observation, and found that both models substantially over-estimated the autoconversion rate, particularly AM4.0 (see their Figure 13b). The Manton and Cotton (1977) scheme and/or its tunings in AM4.0 exaggerates the autoconversion rate of cloud water to rain, which is further magnified by the diagnostic precipitation treatment. Too strong autoconversion triggers too frequent and too early precipitation initiation (Figures 7 and 8), and contributes to insufficient cloudiness (Figures 1 and 3). We compare the cloudiness between AM4-MG2-MCAuto and AM4-MG2, and find that the total and low-level cloud fractions from the CALIPSO simulator are reduced by about 2%.

4. Cloud Radiative Effects

Shortwave and longwave CREs (SWCRE, LWCRE) are among the most important metrics to assess a GCM's performance. Figure 9 exhibits the SWCRE and LWCRE at TOA from the CERES-EBAF observation (Ed4.2), and the differences between the observation and models. In our analysis, positive values indicate absorption. A positive bias means too much absorption. For the SWCRE, the positive bias is the result of a longstanding deficiency of stratocumulus clouds and while not eliminated, the bias is reduced with the use of the MG2 cloud microphysics in AM4-MG2 (Figures 9c and 9e). This echoes the deficient stratocumulus over there, and this deficiency gets improved when MG2 cloud microphysics is effective (Figures 3e and 3h). Negative biases become apparent over the sub-Saharan Africa, western Pacific, and subtropical cumulus regions. Globally the modeled SWCREs are too negative by about 3 W m^{-2}, compared to the observed mean.

The too (negative) strong SWCRE could be associated with low-level clouds that are overly reflective, considering that neither AM4.0 nor AM4-MG2 produce the observed amount of low-level clouds (Section 3.1.2). There is some evidence that a few low clouds are too bright. We therefore now examine their reflectivity. Figure 10a presents the relationship between the albedo at TOA and the non-overlapped low-level cloud fraction, where the overlying mid- and high-level cloud fractions are <5%. The albedo based on the CERES-EBAF observation increases monotonically with the low-level cloud fraction. This is not surprising because more cloud cover is conducive to more radiation reflection. This monotonic increase is captured by both models. AM4.0 and AM4-MG2 albedos are higher than the observation for cloud fraction >20%. The modeled clouds appear to be more reflective or overly bright. As corroborated by Figures 8a–8d and 8g, both models show more frequent occurrences of clouds with thicker optical depth. The thicker clouds are brighter, and can reflect more incoming solar

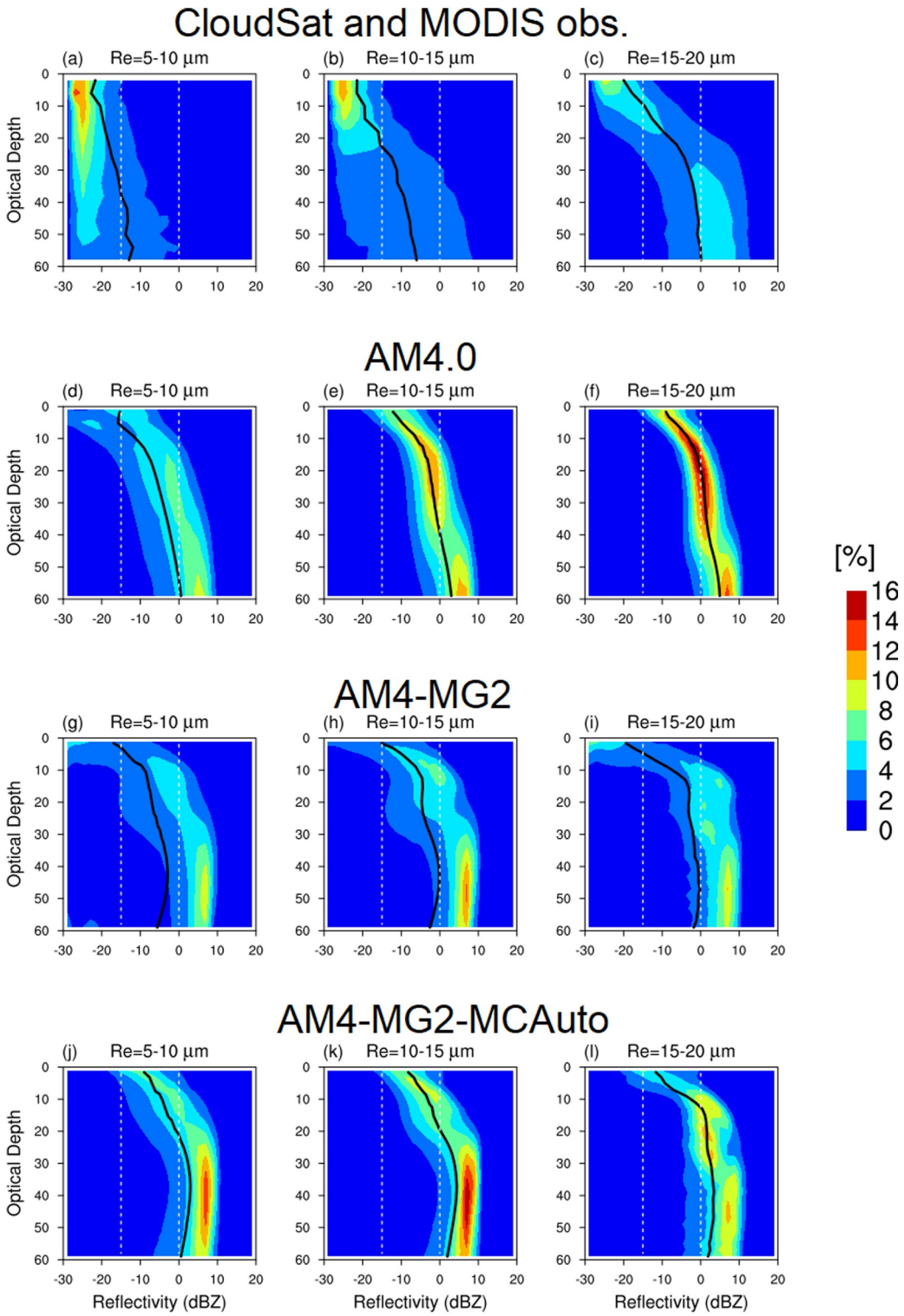

Figure 8. Radar reflectivity probability with respect to in-cloud optical depth in the form of Contoured Frequency by Optical Depth Diagram for different cloud top effective radius: R_e at 5–10 µm, 10–15 µm, and 15–20 µm (from left to right). The observed R_e and optical depth are from MODIS, and the observed reflectivity is provided by CloudSat radar in (a)–(c). The AM4.0, AM4-MG2, and AM4-MG2-MCAuto simulator results are displayed in (d)–(l). The black curve shows the mean relationship between radar reflectivity and optical depth. The white vertical lines separate non-precipitating, drizzling, and precipitating cloud categories. The radar reflectivity probability has been normalized by the in-cloud optical depth.

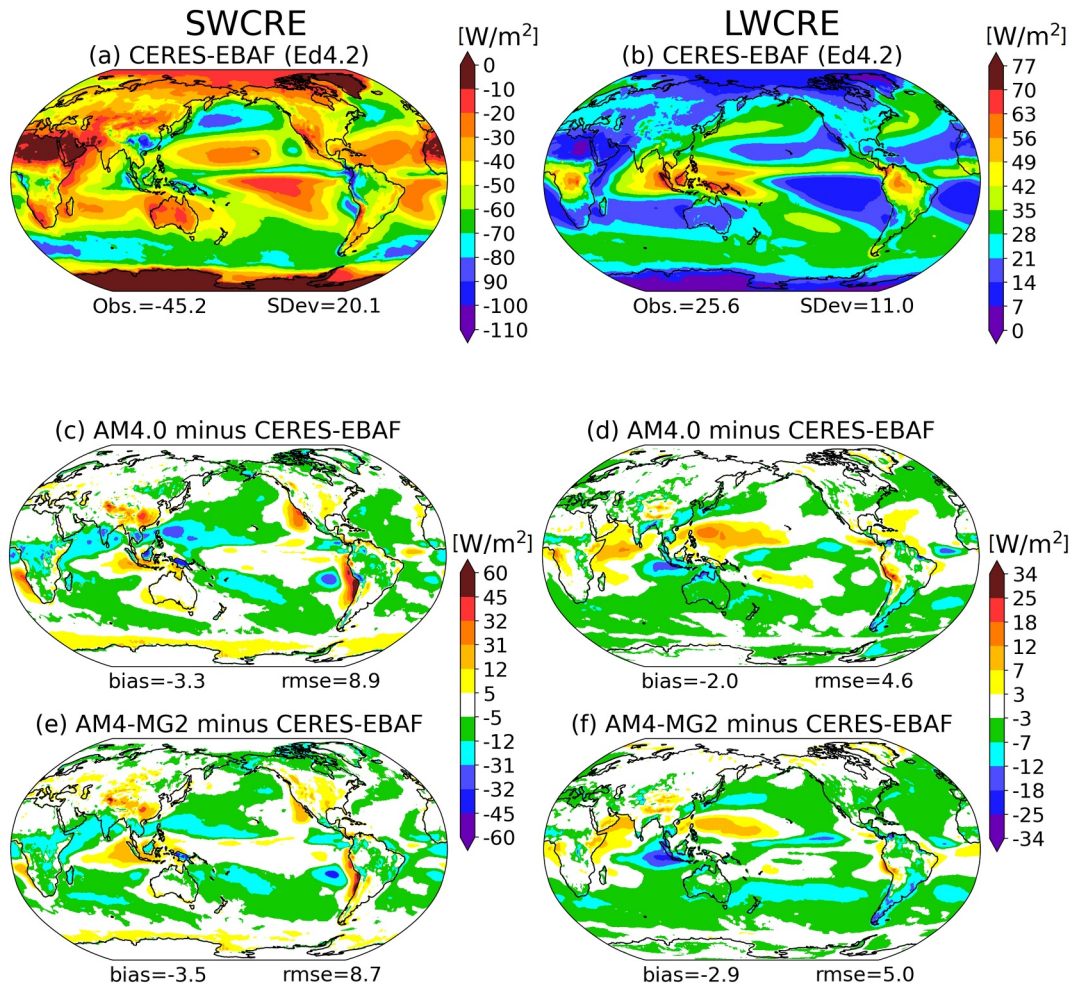

Figure 9. Annually mean SWCRE and LWCRE at TOA from CERES–EBAF Ed4.2 in (a) and (b), from the difference between AM4.0 and observation in (c) and (d), and from the difference between AM4-MG2 and observation in (e) and (f).

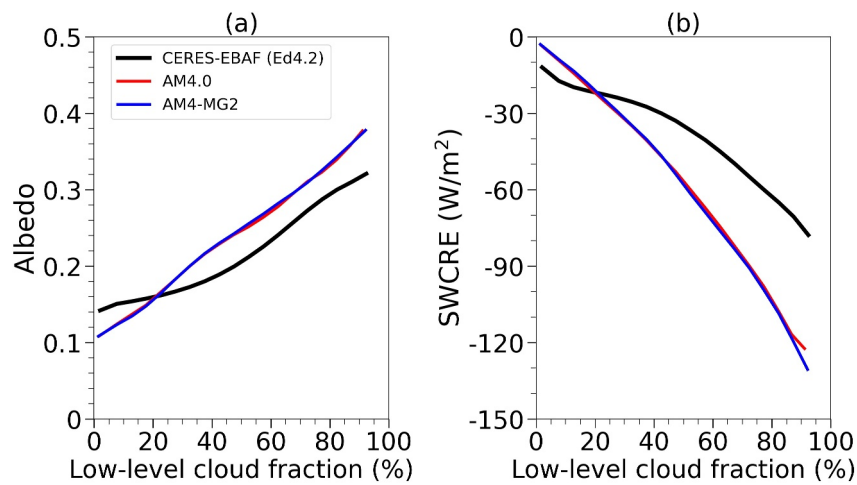

Figure 10. Albedo in (a) and SWCRE in (b) versus non-overlapped low-level cloud fraction over the ocean from 60°S to 60°N.

radiation back to space, and consequently exert stronger (or more negative) SWCRE (Figure 10b). Therefore, the lack of the low-level cloudiness compensates the excessive reflectivity, which yields a relatively low bias of SWCRE.

Similarly, Figures 9b–9d and 9f display the results for the LWCRE. Both models perform reasonably well with relatively small RMSEs of about 5.0 W m^{-2}, although both underestimate the LWCRE compared to CERES-EBAF. At first glance, their bias patterns are qualitatively similar. Negative biases are prevalent over most of the mid-latitude marine regions. Positive biases are seen over the western Pacific and Indian Ocean where negative SWCRE biases occur. This is partly related to the overestimated optically thick high-level clouds (see Figure S2 in Supporting Information S1). These clouds trap excessive longwave but reflect too much shortwave radiation. A quantitative comparison shows that AM4-MG2 has a weaker LWCRE than AM4.0. Their global averages are 22.7 W m^{-2} and 23.6 W m^{-2}, respectively. This seems to be in contrast that AM4-MG2 has more high-level cloudiness than AM4.0 (Figures 4f and 4i). One reason is that AM4-MG2 is more humid, largely because its precipitation efficiency is lower (Guo et al., 2021). More vapor absorbs longwave radiation from the surface and lower atmosphere, but emits at higher altitude and colder temperature. This diminishes the clear-sky outgoing longwave radiation (or the potential longwave radiation that could be trapped by clouds), and results in less LWCRE. Another reason is the ice crystal size. AM4.0 uses a temperature dependent look-up table to diagnose the ice crystal effective radius (Donner et al., 1997; Heymsfield & Platt, 1984). The diagnosed ice effective radius in AM4.0 lies between 10 and 50 μm, and peaks about 50 μm. The ice effective radius in AM4-MG2 is calculated based on the ice mass and number concentrations, and the assumed particle size distribution. It varies from 20 μm to 200 μm, with 2 peaks of around 50 μm and 150 μm (see Figure 10 in Guo et al. (2021)). AM4-MG2 tends to have larger ice crystals. For a given ice mass concentration, AM4-MG2 produces a weaker LWCRE than AM4.0.

5. Summary

The fidelity of cloud simulations is assessed using the satellite simulator package COSP. COSP is run within the host GCMs of AM4.0 and AM4-MG2, and diagnoses cloud properties by emulating satellite observational and retrieval processes. Based on the multiple year AMIP simulations, we evaluate the modeled cloud fraction distribution, vertical profile, cloud water phase partitioning, cloud-to-precipitation transition, and cloud radiative effect. Compared to the up-to-date satellite observations including MODIS, ISCCP, CALIPSO, MISR, CloudSat, and CERES-EBAF, both AM4.0 and AM4-MG2 perform reasonably well. They are able to capture cloud geographical distribution, even though there exist substantial discrepancies across observations particularly for the shallow and/or broken clouds. Both models underestimate the total and low-level cloud fractions. The underestimate is ameliorated in AM4-MG2, partly because the MG2 cloud microphysics enhance the marine stratocumulus clouds along the west coasts of South America, North America, Australia, and Africa. The shortage of stratocumulus is a robust bias in the GFDL GCMs, and many other latest GCMs as well. The enhanced stratocumulus clouds in AM4-MG2 reflects more solar radiation back to space, and thus absorbs less shortwave radiation and reduces the positive bias in SWCRE. Because there exist remarkable discrepancies across the observations, both AM4.0 and AM4-MG2 underestimate midlevel clouds when compared to the ISCCP and CALIPSO observations, but overestimate them relative to the MODIS observations. Although both models underestimate high-level clouds, especially over mid-latitudes, based on the ISCCP observations, they simulate comparable high-level cloud fraction with the MODIS and MISR observations.

Both models successfully capture the co-existence of liquid and ice clouds for temperature of [−40°C, 0°C]. They are able to reproduce the relationship between the ice percentage and temperature, although both under-predict liquid and ice cloud fractions relative to CALIPSO GOCCP. The under-prediction of cloudiness is associated with too frequent and too early precipitation formation. The CloudSat observations show that most clouds do not precipitate (with the non-precipitating cloud occurrence of more than 50%), while both AM4.0 and AM4-MG2 simulate the dominant drizzling clouds. The modeled drizzling occurrence is ∼50%, about 20% higher than observed. The modeled precipitating frequency is ∼20%, about 5% higher than the observation. The CloudSat and MODIS observations reveal that the onset of precipitation is not efficient until the cloud top effective radius exceeds 15 μm, but models could initiate precipitation as early as the cloud top effective radius between the range of 5–10 μm. A sensitivity test highlights that the autoconversion parameterization is of significance for

precipitation initiation. The Seifert and Beheng (2001) scheme in AM4-MG2 provides more realistic auto-conversion rate and thus better represents the cloud-to-precipitation transition.

Both AM4.0 and AM4-MG2 produce too few optically thin and intermediate low-level clouds of $\tau < 23$, but too many optically thick clouds of $\tau > 23$. Their error patterns generally resemble each other. Further cloud reflectivity analysis reveals that the modeled clouds are too bright. For the cloud fraction greater than 20%, the albedo at TOA and SWCRE from either AM4.0 or AM4-MG2 exceed the CERES-EBAF observation in magnitude. Although there is shortage of cloudiness, excessive cloud reflectivity compensates it and keeps the radiation in balance. This implies that a model's capability of reproducing the global radiative fluxes may not explain why the simulated cloudiness and its geographical distribution deviate from observations. Using both radiative fluxes and cloud fraction as well as optical properties offer more consistent constraints for the model climatology. So satellite-based assessment will benefit the model development in GFDL as well as other model centers (Medeiros et al., 2023; Zhang et al., 2024), since the methodologies used in this paper are applicable to other models. The radiation balance provides "top-down" constraints for model climate. Additionally, process level constraints, such as autoconversion and accretion rates, provide valuable "bottom-up" understandings of cloud and precipitation processes. These "bottom-up" constraints need to be taken into account for future model development and evaluation (Held et al., 2019; Suzuki et al., 2013).

Appendix A: Table A1 Provides a List of Acronyms and Abbreviations

Table A1
Table of Acronyms and Abbreviations

AM2	The second generation atmosphere model
AM4.0	The Fourth Generation Atmosphere Model
AMIP	Atmospheric Model Intercomparison Project
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CAM	Community Atmosphere Model
CERES-EBAF	Clouds and the Earth's Radiant Energy System-Energy Balanced and Filled
CFMIP	Cloud Feedback Model Intercomparison Project
CFODD	Contoured Frequency by Optical Depth Diagram
COSP	CFMIP Observation Simulator Package
CMIP5	Coupled Model Intercomparison Project Phase 5
CMIP6	Coupled Model Intercomparison Project Phase 6
CRE	Cloud Radiative Effect
EAM	Energy Exascale Earth System Model Atmosphere Model
ECS	Equilibrium Climate Sensitivity
FV3	Finite-Volume Cubed-Sphere
GCM	General Circulation Model
GFDL	Geophysical Fluid Dynamics Laboratory
GOCCP	GCM-Oriented CALIPSO Cloud Product
HadISST	Hadley Center Sea Ice and Sea Surface Temperature
IPCC	Intergovernmental Panel on Climate Change
INP	Ice Nucleating Particle
ISCCP	International Satellite Cloud Climatology Project
ITCZ	Intertropical Convergence Zone

Table A1
Continued

AM2	The second generation atmosphere model
MG2	Two-moment Morrison-Gettelman microphysics with prognostic precipitation
LWCRE	Longwave Cloud Radiative Effect
MISR	Multiangle Imaging Spectro-Radiometer
MODIS	Moderate Resolution Imaging Spectroradiometer
RMSE	Root Mean Square Error
SWCRE	Shortwave Cloud Radiative Effect
SO	Southern Ocean
SST	Sea Surface Temperature
TOA	Top-Of-Atmosphere

Data Availability Statement

The MODIS and ISCCP-H satellite data sets are available at Pincus et al. (2023) and Young et al. (2018), respectively. The CALIPSO and MISR satellite data sets can be downloaded from https://climserv.ipsl.polytechnique.fr/cfmip-obs/data/GOCCP_V3.1.4/ and https://climserv.ipsl.polytechnique.fr/cfmip-obs/data/MISR/CMOR/clMISR/. The CERES-EBAF Ed4.2 data is accessible from https://asdc.larc.nasa.gov/project/CERES/CERES_EBAF_Edition4.2. The model source code and parameter settings can be obtained from Guo et al. (2020). The model data can be obtained from Zhao et al. (2018c).

Acknowledgments
We thank Drs. Jing Feng and Zhihong Tan for valuable suggestions and comments on an internal review. We also would like to thank two reviewers who helped us to improve the manuscript. Work at LLNL was performed under the auspices of the US DOE by LLNL under contract No. DE-AC52-07NA27344. Dr. Levi G. Silvers is partially supported by NSF under the award number 2327958. K. Suzuki was supported by MEXT program for the Advanced Studies of Climate Change Projection (SENTAN) (Grant JPMXD0722680395) and JAXA/EarthCARE project.

References

Alexander, M. J., & Dunkerton, T. J. (1999). A spectral parameterization of mean-flow forcing due to breaking gravity waves. *Journal of the Atmospheric Sciences*, *56*(24), 4167–4182. https://doi.org/10.1175/1520-0469(1999)056⟨4167:ASPOMF⟩2.0.CO;2

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J. L., Klein, S. A., et al. (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

Bretherton, C. S., McCaa, J. R., & Grenier, H. (2004). A new parameterization for shallow cumulus convection and its application to marine subtropical cloud-topped boundary layers. part i: Description and 1d results. *Monthly Weather Review*, *132*(4), 864–882. https://doi.org/10.1175/1520-0493(2004)132⟨0864:ANPFSC⟩2.0.CO;2

Cesana, G., & Chepfer, H. (2012). How well do climate models simulate cloud vertical structure? A comparison between CALIPSO-GOCCP satellite observations and CMIP5 models. *Geophysical Research Letters*, *39*(20), L20803. https://doi.org/10.1029/2012GL053153

Cesana, G., & Chepfer, H. (2013). Evaluation of the cloud thermodynamic phase in a climate model using CALIPSO-GOCCP. *Journal of Geophysical Research: Atmospheres*, *118*(14), 7922–7937. https://doi.org/10.1002/jgrd.50376

Chan, M. A., & Comiso, J. C. (2013). Arctic cloud characteristics as derived from MODIS, CALIPSO, and CloudSat. *Journal of Climate*, *26*(10), 3285–3306. https://doi.org/10.1175/JCLI-D-12-00204.1

Charney, J. G., Arakawa, A., Baker, D. J., Bolin, B., Dickinson, R. E., Goody, R. M., et al. (1979). *Carbon Dioxide and Climate: A Scientific Assessment*. U.S. National Academy of Sciences.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J. L., Minnis, P., et al. (2010). The GCM-oriented CALIPSO cloud product (CALIPSO-GOCCP). *Journal of Geophysical Research*, *115*(D4), D00H16. https://doi.org/10.1029/2009JD012251

Diner, D., Braswell, B. H., Davies, R., Gobron, N., Hu, J., Jin, Y., et al. (2005). The value of multiangle measurements for retrieving structurally and radiatively consistent properties of clouds, aerosols, and surfaces. *Remote Sensing of Environment*, *97*(4), 495–518. https://doi.org/10.1016/j.rse.2005.06.006

Donner, L. J., Seman, C. J., Soden, B. J., Hemler, R. S., Warren, J. C., Strom, J., & Liou, K. N. (1997). Large-scale ice clouds in the GFDL SKYHI general circulation model. *Journal of Geophysical Research*, *102*(D18), 21745–21768. https://doi.org/10.1029/97jd01488

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fan, S., Ginoux, P., Seman, C. J., Silvers, L. G., & Zhao, M. (2019). Toward improved cloud-phase simulation with a mineral dust and temperature-dependent parameterization for ice nucleation in mixed-phase clouds. *Journal of the Atmospheric Sciences*, *76*(11), 3655–3667. https://doi.org/10.1175/JAS-D-18-0287.1

Fan, S., Knopf, D. A., Heymsfield, A. J., & Donner, L. J. (2017). Modeling of aircraft measurements of ice crystal concentration in the Arctic and a parameterization for mixed-phase cloud. *Journal of the Atmospheric Sciences*, *74*(11), 3799–3814. https://doi.org/10.1175/JAS-D-17-0037.1

Garner, S. T. (2018). Ground-truth model evaluation of subgrid orographic base-flux parameterization. *Journal of the Atmospheric Sciences*, *75*(10), 3653–3670. https://doi.org/10.1175/jas-d-17-0368.1

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., et al. (1999). An overview of the results of the atmospheric model inter-comparison Project (AMIP I). *Bulletin America Meteorology Social*, *80*(1), 29–55. https://doi.org/10.1175/1520-0477(1999)080<0029:aootro>2.0.co;2

Gettelman, A., & Morrison, H. (2015). Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes. *Journal of Climate*, *28*(3), 1268–1287. https://doi.org/10.1175/JCLI-D-14-00103.1

Gettelman, A., Morrison, H., Santos, S., Bogenshutz, P., & Caldwell, P. M. (2015). Advanced two-moment bulk microphysics for global models. Part II: Global model solutions and aerosol-cloud interactions. *Journal of Climate*, *28*(3), 1288–1306. https://doi.org/10.1175/JCLI-D-14-00102.1

Golaz, J.-C., Salzmann, M., Donner, L. J., Horowitz, L. W., Ming, Y., & Zhao, M. (2011). Sensitivity of the aerosol indirect effect to subgrid variability in the cloud parameterization of the gfdl atmosphere general circulation model AM3. *Journal of Climate*, *24*(13), 3145–3160. https://doi.org/10.1175/2010JCLI3945.1

Guo, H., Ming, Y., Fan, S., Zhou, L., Harris, L., & Zhao, M. (2021). Two-moment bulk cloud microphysics with prognostic precipitation in GFDL's atmosphere model AM4.0: Configuration and performance. *Journal of Advances in Modeling Earth Systems*, *13*(6), e2020MS002453. https://doi.org/10.1029/2020MS002453

Guo, H., Ming, Y., Fan, S., Zhou, L., Harris, L., Zhao, M., et al. (2020). NOAA-GFDL/AM4: MG cloud microphysics update with FMS 2020.02 [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.4313356

Harris, L., Zhou, L., Chen, X., & Chen, J.-H. (2020). *The GFDL Finite-Volume Cubed-Sphere dynamical Core,* NOAA Technical Memorandum. OAR GFDL. https://doi.org/10.25923/7h88-c534

Haynes, J. M., & Stephens, G. L. (2007). Tropical oceanic cloudiness and the incidence of precipitation: Early results from cloudsat. *Geophysical Research Letters*, *34*(9), L09811. https://doi.org/10.1029/2007GL029335

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727. https://doi.org/10.1029/2019MS001829

Heymsfield, A. J., & Platt, C. M. R. (1984). A parameterization of the particle size spectrum of ice clouds in terms of the ambient temperature and the ice water content. *Journal of the Atmospheric Sciences*, *41*(5), 846–855. https://doi.org/10.1175/1520-0469(1984)041⟨0846:APOTPS⟩2.0.CO;2

Hillman, B. R., Marchand, R. T., Ackerman, T. P., Mace, G. G., & Benson, S. (2017). Assessing the accuracy of MISR and MISR-simulated cloud top heights using CloudSat- and CALIPSO-retrieved hydrometeor profiles. *Journal of Geophysical Research: Atmospheres*, *122*(5), 2878–2897. https://doi.org/10.1002/2016jd025510

IPCC. (2013). Climate change 2013: The physical science basis. In T. F. Stocker (Ed.), *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

IPCC. (2022). Climate change 2022: Impacts, adaptation, and vulnerability. In H.-O. Pörtner, D. Roberts, M. Tignor, E. Poloczanska, K. Mintenbeck, et al. (Eds.), *Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://doi.org/10.1017/9781009325844

Jing, X., Suzuki, K., Guo, H., Goto, D., Ogura, T., Koshiro, T., & Mülmenstädt, J. (2017). A multimodel study on warm precipitation biases in global models compared to satellite observations. *Journal of Geophysical Research: Atmospheres*, *122*(21), 11806–11824. https://doi.org/10.1002/2017JD027310

Kay, J. E., Bourdages, L., Miller, N. B., Morrison, A., Yettella, V., Chepfer, H., & Eaton, B. (2016). Evaluating and improving cloud phase in the Community Atmosphere Model version 5 using spaceborne lidar observations. *Journal of Geophysical Research: Atmospheres*, *121*(8), 4162–4176. https://doi.org/10.1002/2015JD024699

Kay, J. E., Hillman, B., Klein, S., Zhang, Y., Medeiros, B., Gettelman, A., et al. (2012). Exposing global cloud biases in the Community Atmosphere Model (CAM) using satellite observations and their corresponding instrument simulators. *Journal of Climate*, *25*(15), 5190–5207. https://doi.org/10.1175/JCLI-D-11-00469.1

Klein, S. A., & Jakob, C. (1999). Validation and sensitivities of frontal clouds simulated by the ECMWF model. *Monthly Weather Review*, *127*(10), 2514–2531. https://doi.org/10.1175/1520-0493(1999)127<2514:vasofc>2.0.co;2

Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., & Gleckler, P. J. (2013). Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator. *Journal of Geophysical Research: Atmospheres*, *118*(3), 1329–1342. https://doi.org/10.1002/jgrd.50141

Kristjansson, J. E., Edwards, J. M., & Mitchell, D. L. (2000). Impact of a new scheme for optical properties of ice crystals on climates of two GCMs. *Journal of Geophysical Research*, *105*(D8), 10063–10079. https://doi.org/10.1029/2000jd900015

Kubar, T. L., Hartmann, D. L., & Wood, R. (2009). Understanding the importance of microphysics and macrophysics for warm rain in marine low clouds. Part I: Satellite observations. *Journal of the Atmospheric Sciences*, *66*(10), 2953–2972. https://doi.org/10.1175/2009JAS3071.1

Liu, X., He, T., Liang, S., Li, R., Xiao, X., Ma, R., & Ma, Y. (2023). A monthly 1° resolution dataset of daytime cloud fraction over the Arctic during 2000–2020 based on multiple satellite products. *Earth System Science Data*, *15*(8), 3641–3671. https://doi.org/10.5194/essd-15-3641-2023

Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M., & Smith, R. N. B. (2000). A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Monthly Weather Review*, *128*(9), 3187–3199. https://doi.org/10.1175/1520-0493(2000)128<3187:anblms>2.0.co;2

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., et al. (2018). Clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918. https://doi.org/10.1175/jcli-d-17-0208.1

Manton, M. J., & Cotton, W. R. (1977). Formulation of approximate equations for modeling moist deep convection on the mesoscale. In *Atmospheric Science Paper No. 266*. Colorado State University.

Marchand, R., Ackerman, T., Smyth, M., & Rossow, W. B. (2010). A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS. *Journal of Geophysical Research*, *115*(D16), D16206. https://doi.org/10.1029/2009JD013422

Marchand, R., Mace, G. G., Ackerman, T., & Stephens, G. (2008). Hydrometeor detection using CloudSat—An Earth-orbiting 94-GHz cloud radar. *Journal of Atmospheric and Oceanic Technology*, *25*(4), 519–533. https://doi.org/10.1175/2007JTECHA1006.1

Medeiros, B., Shaw, J., Kay, J. E., & Davis, I. (2023). Assessing clouds using satellite observations through three generations of global atmosphere models. *Earth and Space Science*, *10*(7), e2023EA002918. https://doi.org/10.1029/2023EA002918

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., et al. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth System Models. *Science Advances*, *6*(26), eaba1981. https://doi.org/10.1126/sciadv.aba1981

Meyers, M. P., DeMott, P. J., & Cotton, W. R. (1992). New primary ice nucleation parameterizations in an explicit cloud model. *Journal of Applied Meteorology*, *32*(7), 708–721. https://doi.org/10.1175/1520-0450(1992)031<0708:npinpi>2.0.co;2

Ming, Y., Ramaswamy, V., Donner, L. J., Phillips, V. T. J., Klein, S. A., Ginoux, P. A., & Horowitz, L. W. (2007). Modeling the interactions between aerosols and liquid water clouds with a self-consistent cloud scheme in a General Circulation Model. *Journal of the Atmospheric Sciences*, *64*(4), 1189–1209. https://doi.org/10.1175/JAS3874.1

Morrison, H., & Gettelman, A. (2008). A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, version 3 (CAM3). Part I: Description and numerical tests. *Journal of Climate*, *21*(15), 3642–3659. https://doi.org/10.1175/2008JCLI2105.1

Mueller, K. J., & Coauthors (2017). Assessment of MISR cloud motion vectors (CMVs) relative to GOES and MODIS atmospheric motion vectors (AMVs). *Journal of Applied Meteorology and Climatology*, *56*(3), 555–572. https://doi.org/10.1175/JAMC-D-16-0112.1

Nakajima, T. Y., Suzuki, K., & Stephens, G. L. (2010). Droplet growth in warm water clouds observed by the A-train. Part II: A multi-sensor view. *Journal of the Atmospheric Sciences*, *67*(6), 1897–1907. https://doi.org/10.1175/2010JAS3276.1

Nam, C., Bony, S., Dufresne, J. L., & Chepfer, H. (2012). The "too few, too bright" tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*, *39*(21), L21–L801. https://doi.org/10.1029/2012gl053421

Norris, J. R., & Evan, A. T. (2015). Empirical removal of artifacts from the ISCCP and PATMOS-x satellite cloud records. *Journal of Atmospheric and Oceanic Technology*, *32*(4), 691–702. https://doi.org/10.1175/jtech-d-14-00058.1

Paynter, D. J., & Ramaswamy, V. (2014). Investigating the impact of the shortwave water vapor continuum upon climate simulations using GFDL global models. *Journal of Geophysical Research: Atmospheres*, *119*(18), 10720–10737. https://doi.org/10.1002/2014JD021881

Pincus, R., Hubanks, P. A., Platnick, S., Meyer, K., Holz, R. E., Botambekov, D., & Wall, C. J. (2023). Updated observations of clouds by MODIS for global model assessment. *Earth System Science Data*, *15*(6), 2483–2497. https://doi.org/10.5194/essd-15-2483-2023

Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., & Patrick Hofmann, R. J. (2012). Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. *Journal of Climate*, *25*(13), 4699–4720. https://doi.org/10.1175/JCLI-D-11-00267.1

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C., & Frey, R. A. (2003). The MODIS cloud products: Algorithms and examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 459–473. https://doi.org/10.1109/TGRS.2002.808301

Platnick, S., Li, J. Y., King, M. D., Gerber, H., & Hobbs, P. V. (2001). A solar reflectance method for retrieving the optical thickness and droplet size of liquid water clouds over snow and ice surfaces. *Journal of Geophysical Research*, *106*(D14), 15185–15199. https://doi.org/10.1029/2000jd900441

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002jd002670

Rossow, W. B., Knapp, K. R., & Young, A. H. (2022). International satellite cloud climatology Project: Extending the record. *Journal of Climate*, *35*(1), 141–158. https://doi.org/10.1175/JCLI-D-21-0157.1

Rossow, W. B., & Schiffer, R. A. (1999). Advances in understanding clouds from ISCCP. *Bulletin America Meteorology Social*, *80*(11), 2261–2288. https://doi.org/10.1175/1520-0477(1999)080(2261:AIUCFI)2.0.co;2

Rotstayn, L. D. (1997). A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I, Description and evaluation of the microphysical processes. *Quarterly Journal of the Royal Meteorological Society*, *123*(541), 1227–1282. https://doi.org/10.1256/smsqj.54105

Schwarzkopf, D. M., & Fels, S. (1991). The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *Journal of Geophysical Research*, *96*(D5), 9075–9096. https://doi.org/10.1029/89jd01598

Seifert, A., & Beheng, K. D. (2001). A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmospheric Research*, *59–60*, 265–281. https://doi.org/10.1016/s0169-8095(01)00126-0

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), e2019RG000678. https://doi.org/10.1029/2019RG000678

Stephens, G., & Kummerow, C. D. (2007). The remote sensing of clouds and precipitation from space. *Journal of the Atmospheric Sciences*, *64*(11), 3742–3765. https://doi.org/10.1175/2006jas2375.1

Stephens, G., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., et al., the CloudSat Science Team. (2002). The CloudSat mission and the A-Train. *Bulletin America Meteorology Social*, *83*(12), 1771–1790. https://doi.org/10.1175/bams-83-12-1771

Stubenrauch, C. J., Rossow, W. B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., et al. (2013). Assessment of global cloud datasets from satellites: Project and database initiated by the gewex radiation panel. *Bulletin of the American Meteorological Society*, *94*(7), 1031–1049. https://doi.org/10.1175/bams-d-12-00117

Suzuki, K., Golaz, J. C., & Stephens, G. L. (2013). Evaluating cloud tuning in a climate model with satellite observations. *Geophysical Research Letters*, *40*(16), 4464–4468. https://doi.org/10.1002/grl.50874

Suzuki, K., Nakajima, T. Y., & Stephens, G. L. (2010). Particle growth and drop collection efficiency of warm clouds as inferred from joint CloudSat and MODIS observations. *Journal of the Atmospheric Sciences*, *67*(9), 3019–3032. https://doi.org/10.1175/2010JAS3463.1

Suzuki, K., Stephens, G. L., Bodas-Salcedo, A., Wang, M., Golaz, J.-C., Yokohata, T., & Tsuyoshi, K. (2015). Evaluation of the warm rain formation process in global models with satellite observations. *Journal of the Atmospheric Sciences*, *72*(10), 3996–4014. https://doi.org/10.1175/JAS-D-14-0265.1

Swales, D. J., Pincus, R., & Bodas-Salcedo, A. (2018). The cloud feedback model Intercomparison Project observational simulator package: Version 2. *Geoscientific Model Development*, *11*(1), 77–81. https://doi.org/10.5194/gmd-11-77-2018

Tiedtke, M. (1993). Representation of clouds in large-scale models. *Monthly Weather Review*, *121*(121), 3040–3061. https://doi.org/10.1175/1520-0493(1993)121<3040:rocils>2.0.co;2

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., et al. (2009). Overview of the CALIPSO mission and CALIOP data processing algorithms. *Journal of Atmospheric and Oceanic Technology*, *26*(11), 2310–2323. https://doi.org/10.1175/2009jtecha1281.1

Wood, R., Kubar, T. L., & Hartmann, D. L. (2009). Understand-ing the importance of microphysics and macrophysics for warm rain in marine low clouds. Part II: Heuristic mod-els of rain formation. *Journal of the Atmospheric Sciences*, *66*(10), 2973–2990. https://doi.org/10.1175/2009JAS3072.1

Young, A. H., Knapp, K. R., Inamdar, A., Hankins, W., & Rossow, W. B. (2018). The international satellite cloud climatology Project H-series climate data record product. *Earth System Science Data*, *10*(1), 583–593. https://doi.org/10.5194/essd-10-583-2018

Zelinka, M. D., Chao, L.-W., Myers, T. A., Qin, Y., & Klein, S. A. (2025). Technical note: Recommendations for diagnosing cloud feedbacks and rapid cloud adjustments using cloud radiative kernels. *Atmospheric Chemistry and Physics*, *25*(3), 1477–1495. https://doi.org/10.5194/acp-25-1477-2025

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in cmip6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782. https://doi.org/10.1029/2019GL085782

Zhang, Y., Xie, S., Covey, C., Lucas, D. D., Gleckler, P., Klein, S. A., et al. (2012). Regional assessment of the parameter-dependent performance of CAM4 in simulating tropical clouds. *Geophysical Research Letters*, *39*(14), L14708. https://doi.org/10.1029/2012GL052184

Zhang, Y., Xie, S., Lin, W., Klein, S. A., Zelinka, M., Ma, P.-L., et al. (2019). Evaluation of clouds in version 1 of the E3SM atmosphere model with satellite simulators. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1253–1268. https://doi.org/10.1029/2018MS001562

Zhang, Y., Xie, S., Qin, Y., Lin, W., Golaz, J.-C., Zheng, X., et al. (2024). Understanding changes in cloud simulations from E3SM version 1 to version 2. *Geoscientific Model Development*, *17*, 169–189. https://doi.org/10.5194/gmd-17-169-2024

Zhao, M., Golaz, C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (2018c). NOAA-GFDL GFDL-AM4 model output [Dataset]. *Earth System Grid Federation*. https://doi.org/10.22033/ESGF/CMIP6.1401

Zhao, M., Golaz, J.-C., Held, I., Guo, H., Balaji, V., Benson, R., et al. (2018a). The GFDL global atmosphere and land model AM4. 0/LM4.0- Part I: Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems*, *10*(3), 691–734. https://doi.org/10.1002/2017ms001208

Zhao, M., Golaz, J.-C., Held, I., Guo, H., Balaji, V., Benson, R., et al. (2018b). The GFDL global atmosphere and land model AM4. 0/LM4.0- Part II: Model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, *10*(3), 735–769. https://doi.org/10.1002/2017ms001209

Zhao, M., Golaz, J.-C., Held, I. M., Ramaswamy, V., Lin, S.-J., Ming, Y., et al. (2016). Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics. *Journal of Climate*, *29*(2), 543–560. https://doi.org/10.1175/JCLI-D-15-0191.1