Tropical anvil cirrus are highly sensitive to ice microphysics within a nudged global storm-resolving model

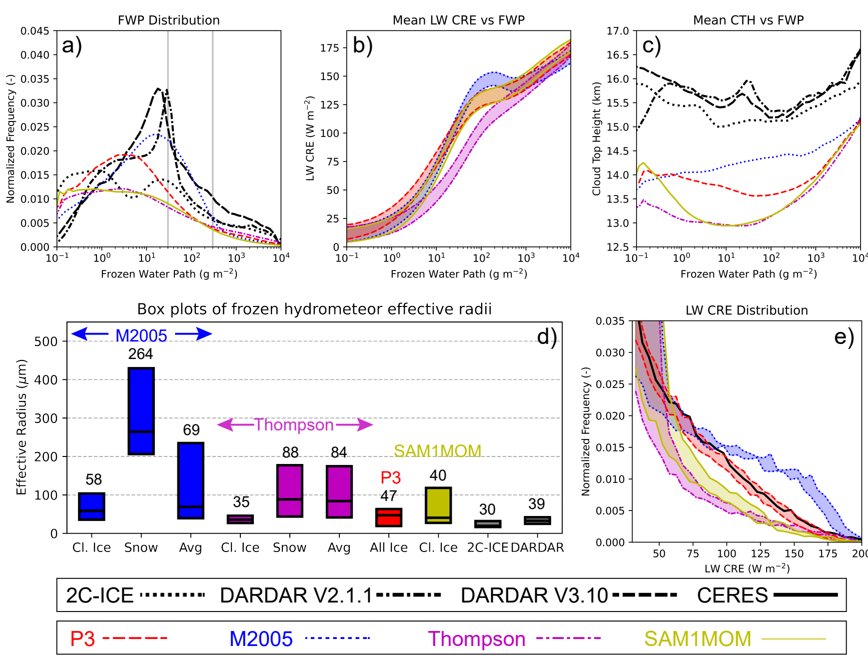
Rachel Atlas[1], Christopher S. Bretherton[2], Adam B Sokol[1], Peter N. Blossey[1], and Marat Khairoutdinov[3]

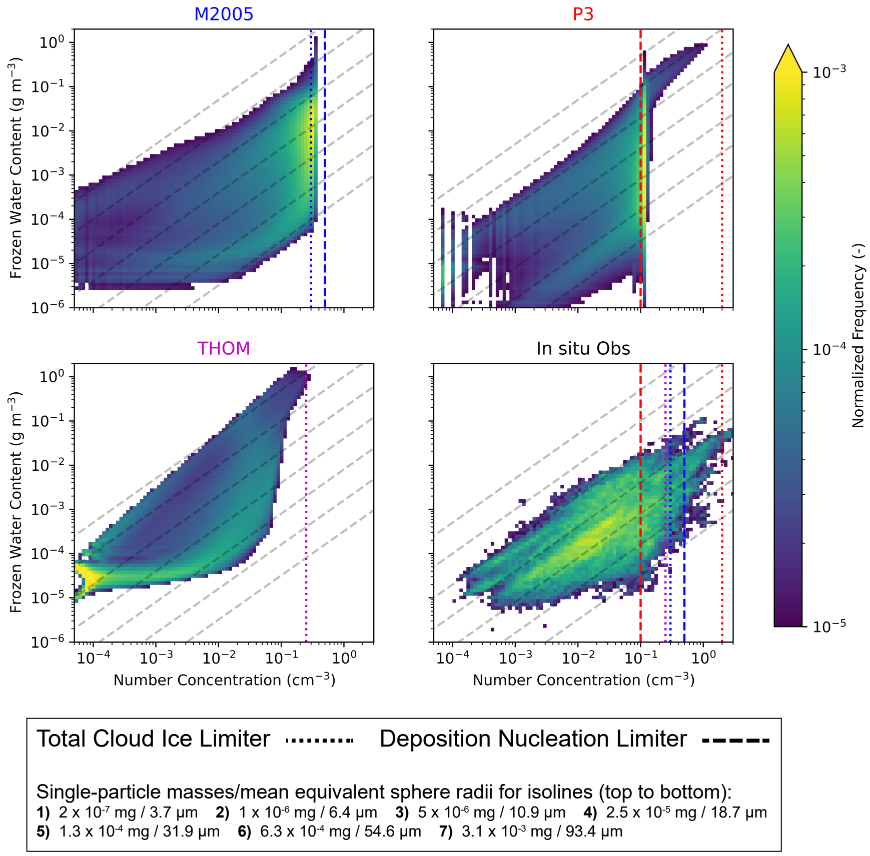
[1]University of Washington
[2]Allen Institute for Artificial Intelligence
[3]Stony Brook University

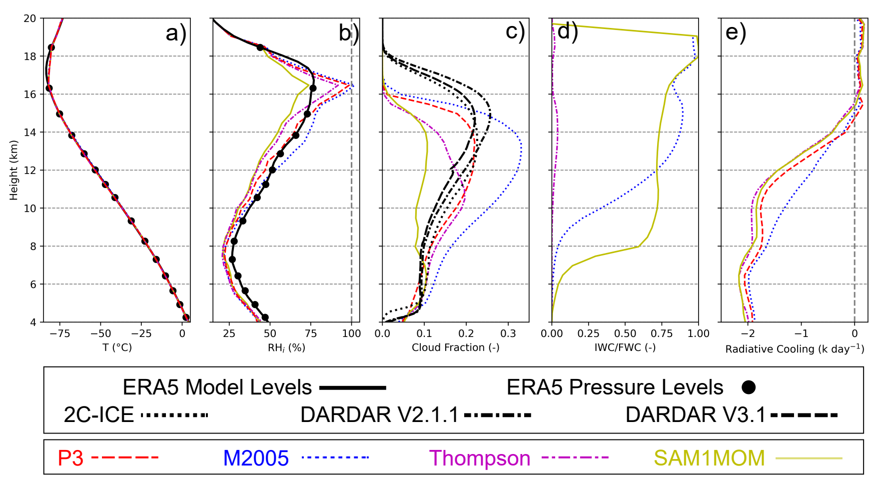
September 11, 2023

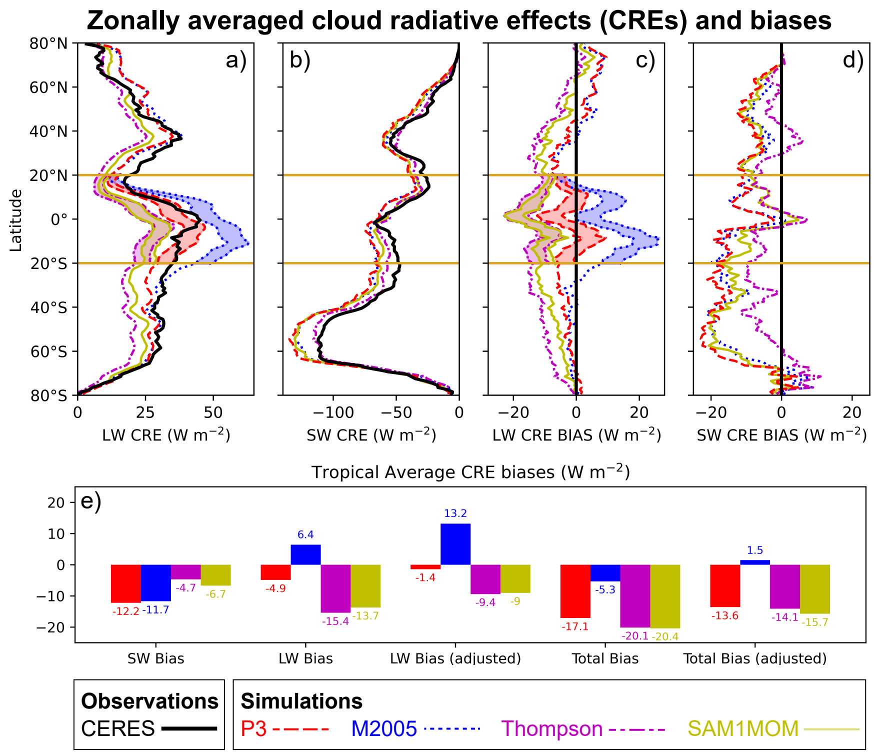
Abstract

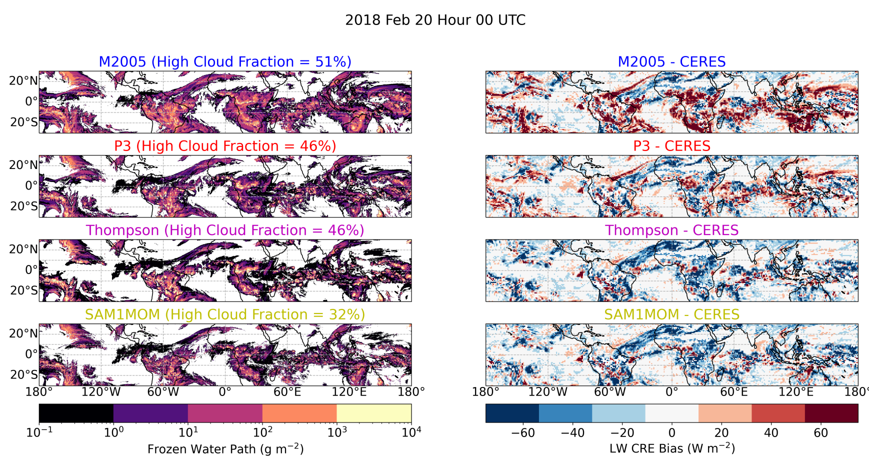
Cirrus dominate the longwave radiative budget of the tropics. For the first time, we quantify the variability in cirrus properties and longwave cloud radiative effects (CREs) that arises from differences in microphysics within nudged global storm-resolving simulations from a single model. Nudging allows us to compute radiative biases precisely using coincident satellite measurements and to fix the large-scale dynamics across our set of simulations and isolate the influence of microphysics. We run five-day simulations with four commonly-used microphysics schemes of varying complexity (SAM1MOM, Thompson, M2005 and P3) and find that the tropical average longwave CRE varies over 20 W m$ˆ{-2}$ between schemes. P3 best reproduces observed longwave CRE. M2005 and P3 simulate cirrus with realistic frozen water path but unrealistically high ice crystal number concentrations which commonly hit limiters and lack the variability and dependence on frozen water content seen in aircraft observations. Thompson and SAM1MOM have too little cirrus.




M2005
P3
THOM
In situ Obs
Frozen Water Content (g m^{-3})
Number Concentration (cm^{-3})
Normalized Frequency (-)
Total Cloud Ice Limiter
Deposition Nucleation Limiter
Single-particle masses/mean equivalent sphere radii for isolines (top to bottom):
1) 2 x 10^{-7} mg / 3.7 µm 2) 1 x 10^{-6} mg / 6.4 µm 3) 5 x 10^{-6} mg / 10.9 µm 4) 2.5 x 10^{-5} mg / 18.7 µm
5) 1.3 x 10^{-4} mg / 31.9 µm 6) 6.3 x 10^{-4} mg / 54.6 µm 7) 3.1 x 10^{-3} mg / 93.4 µm


a)
b)
c)
d)
e)
Height (km)
T (°C)
RH$_i$ (%)
Cloud Fraction (-)
IWC/FWC (-)
Radiative Cooling (k day^{-1})
ERA5 Model Levels
ERA5 Pressure Levels
2C-ICE
DARDAR V2.1.1
DARDAR V3.1
P3
M2005
Thompson
SAM1MOM





Hosted file

970999_0_video_11270450_rz3bd9.gif available at https://authorea.com/users/549349/articles/659477-tropical-anvil-cirrus-are-highly-sensitive-to-ice-microphysics-within-a-nudged-global-storm-resolving-model

Tropical anvil cirrus are highly sensitive to ice microphysics within a nudged global storm-resolving model

R.L. Atlas[1], C.S. Bretherton[2], A.B. Sokol[1], P.N. Blossey[1], M.F. Khairoutdinov[3]

[1]Dept. of Atmospheric Sciences, University of Washington, Seattle, WA
[2]Allen Institute for Artificial Intelligence, Seattle, WA
[3]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY
[1]3920 Okanogan Ln, Seattle, WA 98195
[2]2157 N Northlake Way #110, Seattle, WA 98103
[3]Endeavour, 145, Stony Brook, NY 11790

Key Points:

- Nudged global storm-resolving simulations are valuable for microphysics sensitivity studies.
- Mean tropical longwave cloud radiative effect biases vary over 20 W m^{-2} depending on microphysics scheme.
- Two-moment schemes outperform simpler one-moment and partial double-moment schemes, and P3 has the smallest longwave radiative bias.

Corresponding author: Rachel Atlas, `rachel.atlas@lmd.ipsl.fr`

Abstract

Cirrus dominate the longwave radiative budget of the tropics. For the first time, we quantify the variability in cirrus properties and longwave cloud radiative effects (CREs) that arises from differences in microphysics within nudged global storm-resolving simulations from a single model. Nudging allows us to compute radiative biases precisely using coincident satellite measurements and to fix the large-scale dynamics across our set of simulations and isolate the influence of microphysics. We run five-day simulations with four commonly-used microphysics schemes of varying complexity (SAM1MOM, Thompson, M2005 and P3) and find that the tropical average longwave CRE varies over 20 W m^{-2} between schemes. P3 best reproduces observed longwave CRE. M2005 and P3 simulate cirrus with realistic frozen water path but unrealistically high ice crystal number concentrations which commonly hit limiters and lack the variability and dependence on frozen water content seen in aircraft observations. Thompson and SAM1MOM have too little cirrus.

Plain Language Summary

Recently, advancements in computing have made it possible for atmospheric scientists to simulate Earth's global atmosphere with higher resolution than ever before. This new generation of models, called global-storm resolving models, have a horizontal grid spacing of just a few kilometers, which permits the formation of thunderstorms. As a result, they simulate clouds more realistically than traditionally climate and weather models and are a great tool for diagnosing cloud biases in atmospheric models. Here, we run a single global storm-resolving model with four different representations of cloud physics called M2005, P3, SAM1MOM and Thompson. We evaluate simulated tropical cirrus, which are stratiform ice clouds at the top of the troposphere that reduce the amount of infrared radiation emitted by the Earth, with satellite and aircraft data to see which representations have the best performance. SAM1MOM and Thompson make too little cirrus causing too much infrared radiation to be emitted, M2005 makes too much cirrus, causing too little infrared radiation to be emitted, and P3 makes about the right amount.

1 Introduction

Anvil cirrus, which flow outward from deep convective cores (Deng et al., 2016), absorb longwave radiation from Earth's surface and re-emit it at colder temperatures, thereby reducing outgoing longwave radiation and heating the atmosphere (Hartmann et al., 2001). Differences in their representation in global climate models (GCMs), which stem from diverse model dynamics and physical parameterizations, are a major source of uncertainty in constraining the longwave radiative budget of the tropics and cloud climate feedbacks (Sherwood et al., 2020). Here, we quantify the variability in tropical longwave cloud radiative effect (CRE) that arises from differences in model microphysics across a set of global storm-resolving simulations (GSRMs), and we identify an important avenue for improving ice microphysics and the representation of anvil cirrus.

Anvil cirrus are sensitive to the representation of deep convection and ice microphysics. These influences are difficult to disentangle in most global models, including high resolution GCMs, where both are parameterized. GSRMs, which typically have sub-5 km horizontal grid spacing and explicit rather than parameterized deep convection, provide a unique opportunity to isolate the influence of ice microphysics.

GSRMs are computationally expensive and thus are typically run for short durations ranging from a few days to a year. Comparisons of simulated CREs from short-duration simulations with climatological observations are sensitive to sampling bias. We address this issue by nudging our simulations to reanalysis to prevent the microphysics from feeding back onto the large-scale flow. This approach has many advantages including 1) allowing comparisons with coincident real-world observations, 2) isolating the direct impact of differences in model microphysics on simulated cloud properties, and 3) reducing model spin-up time.

We run our nudged GSRM with four widely used microphysics schemes of varying complexity (single-moment, partial double-moment, and double-moment). We evaluate our simulations with remote sensing observations, including the newly released DARDAR-CLOUD v3.10 dataset (Delanoë & Hogan, 2010), and in-situ observations, leveraging a new dataset aggregating measurements from multiple aircraft campaigns that sampled cirrus clouds (Krämer, Rolf, Spelten, Afchine, et al., 2020).

2 Data

Four five-day simulations are run with the Global System for Atmospheric Modelling (Khairoutdinov et al., 2022). They are set up identically, as described in Atlas et al. (2022), except that they are run with different bulk microphysics schemes: M2005 (Morrison, Curry, & Khvorostyanov, 2005; Morrison et al., 2009), Thompson (Thompson et al., 2008), P3 (Morrison & Milbrandt, 2015) with one ice class, and SAM1MOM (Khairoutdinov & Randall, 2003). All schemes except SAM1MOM were originally developed for the Weather Research & Forecasting model (Skamarock & Klemp, 2008), and are also operational within the Model for Prediction Across Scales (Skamarock et al., 2012) GSRM. The Community Earth System Model (Danabasoglu et al., 2020) and the Energy Exascale Earth System Model (E3SM) (Golaz et al., 2019) GCMs use microphysics schemes closely related to M2005, and the Simple Cloud Resolving E3SM Atmosphere Model GSRM (Caldwell et al., 2021) uses P3 microphysics.

Key differences in the representation of ice processes across the four schemes are summarized in Text S1. The simulations have approximately 4 km horizontal grid spacing in the tropics and about 500 m vertical grid spacing between 5 and 19 km. Deep convection is permitted but somewhat under-resolved using this grid spacing (Bryan et al., 2003). Simulations are initialized from ERA5 reanalysis (Hersbach et al., 2020) at 00 UTC 16 Feb. 2018. We analyze days 2-5 of the simulations (17-20 February 2018) throughout this study, allowing one day for model spinup, long enough for cloud statistics to equilibrate (Atlas et al., 2022). Simulated temperature and horizontal winds (but not humidity or clouds) are nudged to ERA5 reanalysis with a damping timescale of 24 hours.

Simulated longwave and shortwave CREs are compared with coincident retrievals (overlapping the time period of the simulations) from Clouds and the Earth's Radiant Energy System level 3 data (Doelling et al., 2013; NASA/LARC/SD/ASDC, 2017), referred to hereafter as CERES. CERES has hourly temporal resolution and 1° x 1° horizontal resolution.

Retrieved frozen water content (FWC) and effective radii (r_e) from the DARDAR-CLOUD dataset (Delanoë & Hogan, 2010) versions V2.1.0 and V3.10 (Cazenave et al., 2019) and the Cloudsat and CALIPSO Ice Cloud Property Product (2C-ICE) (Deng et al., 2015) version RF05 are used to evaluate simulated anvil cirrus macrophysics. These retrievals have a horizontal resolution of 1.4 km, comparable to that of the simulations.

The vertical resolution of DARDAR and 2C-ICE are 60 m and 240 m, respectively. Because these retrievals are sparse in space and time and direct comparisons cannot be made for the simulated days, we use February data from the years 2007-2012.

Simulated microphysics are evaluated with in situ airborne observations of ice crystal number concentration (N_{ice}) and FWC from five campaigns, which are included in the 'Microphysics Guide to Cirrus' (Krämer, Rolf, & Spelten, 2020), as described in Krämer, Rolf, Spelten, Afchine, et al. (2020). Text S2-S3 and Figures S1-S3 further discuss our use of DARDAR, 2C-ICE and the 'Microphysics Guide to Cirrus'.

3 Microphysics schemes exhibit wide-ranging tropical longwave cloud radiative effects

Figure 1 compares day 2-5 mean simulated CREs with CERES. Throughout this study, radiative fluxes are defined as positive downwards, so that negative CREs indicate energy lost from the Earth. Shortwave CRE biases (panel b) are largest and most scheme-dependent over the Southern Ocean, mainly due to differences in marine boundary layer clouds (Atlas et al., 2022).

In this study, we focus on the region between the horizontal parallel lines at 20°N and 20°S, hereafter referred to as 'the tropics', where longwave CRE is highly sensitive to microphysics (panel a).

Zonally-averaged longwave and shortwave CRE biases for each scheme are plotted on panels c-d. As discussed in Section 4, tropical cloud top heights (CTHs) are biased low in all simulations due to temperature biases around the cold point, causing longwave CREs to be underestimated. Shaded regions show the changes in LW CRE we get when we account for the CTH bias, following Text S3. The lines on the left of the shaded regions are unadjusted LW CRE biases and the lines on the right are adjusted LW CRE biases. Panel e shows area-weighted tropical mean CRE biases for the simulations. LW and SW CRE biases vary over ranges of 22 and 7.5 W m^{-2}, respectively. While all simulations have a bright (negative) tropical shortwave CRE bias, the sign of the longwave CRE bias differs between M2005 and the other schemes, both before and after the adjustment. P3 has the least biased longwave CRE, and is nearly unbiased after the adjustment. M2005 has the smallest total CRE bias, which it achieves through compensating LW and SW biases.

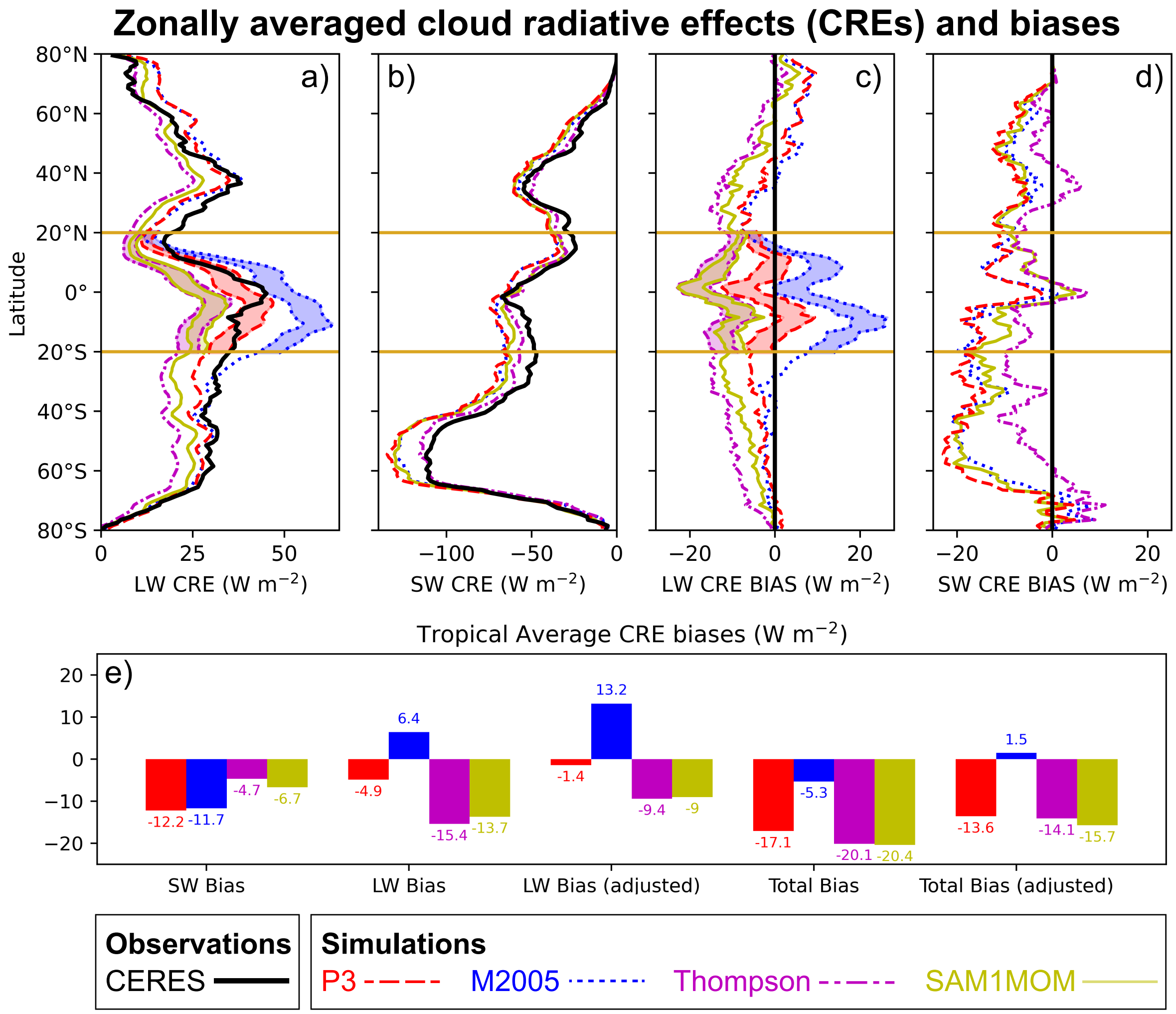


Figure 1. **(a-b)** Zonal average top of atmosphere CREs and **(c-d)** their biases vs. CERES. Horizontal parallel lines delineate the tropical analysis region (20°S - 20°N). Shaded regions show the magnitude of the CTH bias correction **(e)** Tropical average CRE biases for (left to right) SW, LW, adjusted LW, total, and adjusted total

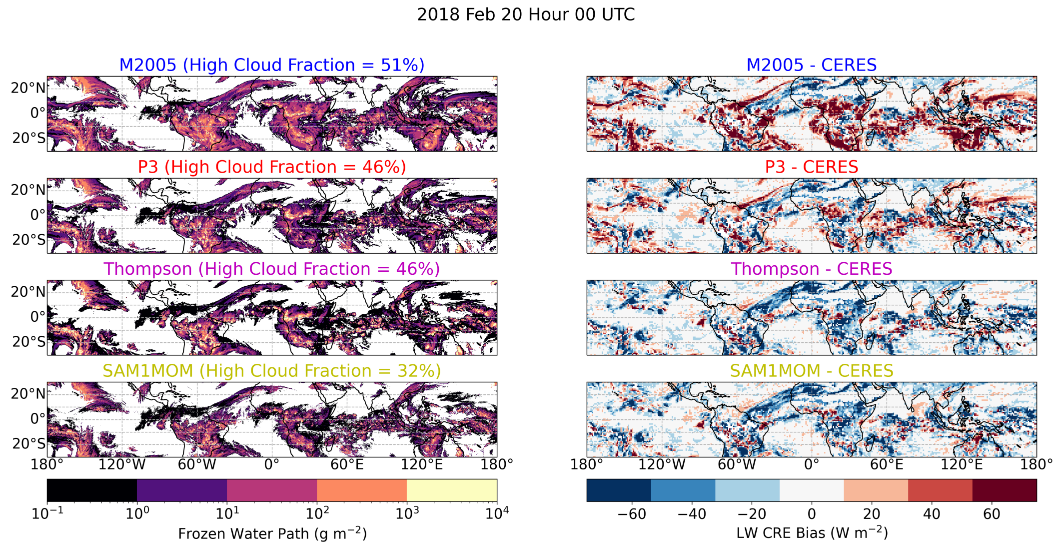


Figure 2. **left:** Snapshots of simulated FWP for columns with CTH $\geq$ 10 km on the simulations' native grid. **right:** Coincident snapshots of longwave CRE bias compared to CERES on a coarsened 1° x 1° grid.

4 Variability in anvil cirrus coverage and optical properties lead to diverse longwave cloud radiative effects

Figure 2 shows coincident snapshots at an arbitrarily chosen time of simulated frozen water path (FWP, the sum of the cloud ice, snow and graupel water paths) for columns containing high cloud, on the left, and biases in simulated longwave CRE, which have not been adjusted to account for the CTH bias, coarsened to a 1° x 1° grid, on the right. Columns with high cloud have a cloud top height (CTH) exceeding 10 km, where CTH is defined as the highest model level with FWC (the sum of the cloud ice, snow and graupel water contents) $\geq 10^{-4}$ g m^{-3} (the limit of lidar detectability as discussed in Text S2). The fraction of columns within the mapped area that meet these criteria is listed in the title of each plot panel. The coarsened longwave CRE bias is sensitive to both cloud fraction and cloud radiative properties. Animation S1 loops through versions of Figure 2 for each of the 96 hours of model output within days 2-5 of the simulations, showing that any hourly snapshot is representative of the entire four day period.

M2005 has the largest high cloud fraction and extensive areas of negative longwave cloud biases, associated with deep convection (FWP $> 10^3$ g m^{-2}) and anvil cirrus (10 $\leq$ FWP $\leq$ 10^3 g m^{-2}). Thompson and SAM1MOM have positive longwave biases in most areas of anvil cirrus. P3 has a mixture of positive and negative biases associated with anvil cirrus, and the fewest areas with large biases of either sign. Adjusting the LW CRE to account for the low bias in CTH reduces LW CRE biases in P3, Thompson and SAM1MOM and increases them in M2005 (Figure S5).

Figure 3 statistically summarizes relationships between high cloud properties and longwave cloud biases, using CERES, DARDAR and 2C-ICE to provide observational constraints on the simulations. The CALIPSO lidar used by DARDAR and 2C-ICE has greater sensitivity at night, during which it can detect FWCs $\geq 10^{-4}$ g m^{-3} (Text S2). Thus, we use DARDAR and 2C-ICE data from the nighttime A-train overpass, which crosses the equator at approximately 1:30 AM local time. For consistency, we also sample CERES and the simulations at night. FWCs $< 10^{-4}$ g m^{-3} are filtered out of the simulations and satellite retrievals.

In Figure 3a, we evaluate distributions of simulated FWP from columns containing high cloud (CTH $\geq$ 10 km) using DARDAR and 2C-ICE. The simulations and the two DARDAR datasets have unimodal distributions of FWP whereas 2C-ICE has a bimodal distribution. The discrepancy between DARDAR and 2C-ICE for FWPs $<$ 30 g m^{-2}, noted by Hong et al. (2016), emphasizes limitations on constraining FWP from CALIPSO in tropical cirrus too thin to be detected by CloudSat. Satellite retrievals from deep convective cores (FWPs $\geq 10^3$ g m^{-2}) are also uncertain (Delanoë & Hogan, 2010). Thus we focus on FWPs between 30 and 300 g m^{-2}, the region bounded by vertical grey lines in Figure 3a, where the retrievals are more certain. In this thickness range, M2005 overestimates cirrus coverage, and SAM1MOM, Thompson, and P3 underestimate it.

Figure 3b shows mean longwave CRE as a function of FWP for the simulations. Shaded regions show the changes in LW CRE we get when we account for the CTH bias. The lines below the shaded region are unadjusted and the lines above the shaded region are adjusted. We do not show an observational comparison here because the retrieved FWP from DARDAR and 2C-ICE is 1D and cannot be matched with the coarsely gridded LW CRE from CERES. M2005 has the strongest longwave CRE for anvil cirrus, and Thompson has the weakest. Variability in longwave CRE for a fixed FWP can be caused by differences in cloud top temperature, which is tightly linked to CTH in the tropics. Figure 3c shows mean CTH as a function of FWP for the simulations and the satellite retrievals. CTH is biased low in all simulations, as explained later in this section. M2005

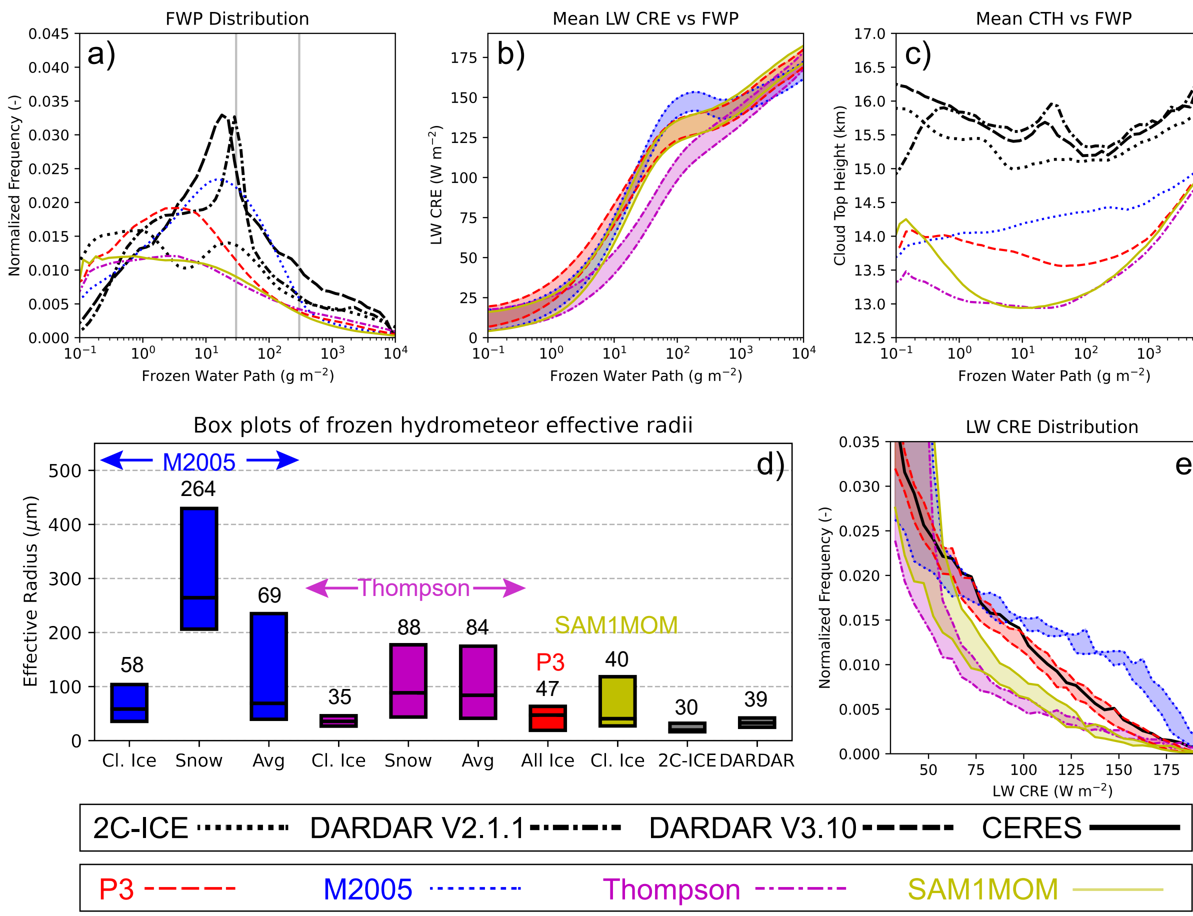


Figure 3. Tropical nighttime: **a)** PDF of FWP **b)** Mean longwave CRE binned by FWP **c)** Mean CTH binned by FWP **d)** Box plots with medians (black lines and numbers printed above each box) and inter-quartile ranges of frozen hydrometeor r_e (for each radiatively-active ice species and the average, calculated as described in the text). Cl. Ice is short for Cloud Ice. **e)** PDF of longwave CRE for 1° x 1° boxes. Only columns with CTH $\geq$ 10 km and grid cells with FWC $\geq 10^{-4}$ g m^{-3} are used in panels **a-d**. Vertical grey lines in panel **a** bound the region where retrievals of FWP are most certain. In panels **b** and **e**, shaded regions show the magnitude of the CTH bias correction, and the area under the curves represents the fraction of high cloud columns and the fraction of 1° x 1° degree boxes with longwave CRE > 25 W m^{-2}, respectively.

has the highest CTHs for FWPs > 1 g m^{-2}, and Thompson and SAM1MOM have the lowest.

Differences between the simulations in Figure 3b could also come from differences in effective radii (r_e). Figure 3d shows box plots of r_e of frozen hydrometeors for the simulations, 2C-ICE and DARDAR V3.10 (the two versions of DARDAR have similar r_e). For M2005 and Thompson, $r_{e,avg}$ is an optical depth preserving average of the cloud ice and snow effective radii, $r_{e,i}$ and $r_{e,s}$, which is directly comparable to satellite-retrieved r_e. For P3, there is only one frozen hydrometeor class and for SAM1MOM, only cloud ice is radiatively active, so the snow contribution to r_e is neglected.

In M2005, the median $r_{e,avg}$ is similar to the median $r_{e,i}$ because cloud ice dominates the frozen hydrometeor mass. In Thompson, the median $r_{e,avg}$ is similar to the median $r_{e,s}$ because snow dominates the frozen hydrometeor mass. This causes Thompson to have an unrealistically large $r_{e,avg}$, which contributes to it having the weakest longwave CRE in Figure 3b. All simulations have larger median r_e than observed, consistent with Stanford et al. (2017).

Figure 3e shows the tail of the histogram of 1° x 1° nighttime longwave CRE for the simulations and CERES, which includes areas that contribute most to the tropical average and to differences between simulations and CERES.

Shaded regions show the changes in LW CRE we get when we account for the CTH bias. The lines below the shaded region are unadjusted and the lines above the shaded region are adjusted. P3 does a strikingly good job at matching the observations. M2005 has too many areas with average longwave CRE > 100 W m^{-2} because it has more anvil cirrus than DARDAR and 2C-ICE (Figure 3a). Thompson and SAM1MOM have too few areas with average longwave CRE > 30 W m^{-2}, due to deficient anvil cirrus and (for Thompson) unrealistically large r_e.

Figure 4 compares simulated vertical profiles of thermodynamic and cloud properties with two ERA5 datasets, DARDAR and 2C-ICE. Figure 4a shows temperature profiles from ERA5 on 37 pressure levels and 137 model levels. In all simulations, temperature was nudged to pressure-level data (black dots), linearly interpolated to the gSAM model levels. The ERA5 model level data (black line) better resolves the 16-18 km layer,

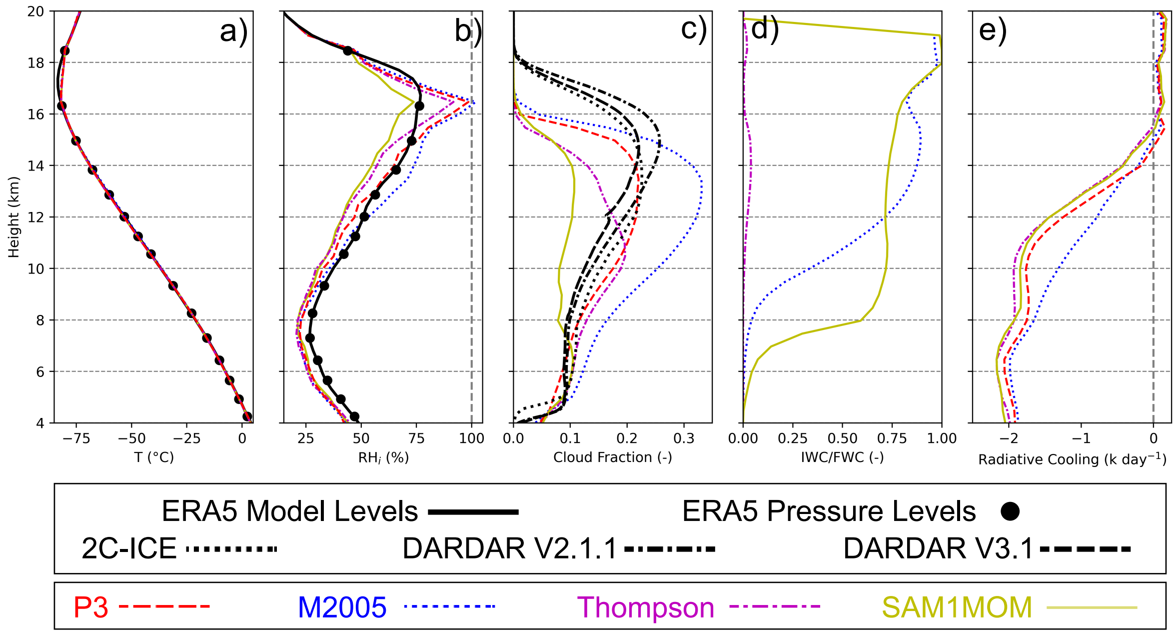


Figure 4. Vertical profiles of tropical nighttime **a)** median temperature, **b)** median RH$_i$, **c)** cloud fraction (FWC $\geq 10^{-4}$ g m^{-3} only), **d)** IWC/FWC, and **e)** mean longwave radiative cooling. P3 has only one ice class so IWC/FWC cannot be computed.

which includes the cold point at 17.3 km. All simulations have a warm bias in that layer and a cold point near 16 km instead of 17.3 km.

Figure 4b show profiles of median relative humidity with respect to ice (RH$_i$). SAM1MOM has a lower median RH$_i$ than the other simulations and ERA5, particularly above 14 km, possibly because it uses saturation adjustment for cloud ice, preventing RH$_i$ from ever exceeding 100%. The importance of representing ice supersaturation for simulated cirrus properties has been noted in previous studies such as Lohmann et al. (2008). The other simulations have higher RH$_i$ than ERA5 near the cold point, but ERA5 may be biased by its internal ice microphysical modeling assumptions in the tropical tropopause layer, where routine observations of the very low water vapor concentration are uncertain.

Figure 4c shows profiles of cloud fraction. For all simulations, the highest cloud tops are 2 km lower than observed, due to their artificially lowered cold point altitudes which result from nudging the model temperature to ERA5 temperature on pressure levels. Below 14.5 km, P3 agrees well with both DARDAR datasets and 2C-ICE, M2005 overestimates cloud fraction, and SAM1MOM and Thompson underestimate it. In M2005 and P3, cloud fraction increases monotonically up to the base of the tropical tropopause layer at 14 km. SAM1MOM has a nearly constant cloud fraction throughout the troposphere. Thompson's peak cloud fraction is only at 10.5 km. Figure 4d, which shows the average mass of IWC (cloud ice only) divided by FWC (cloud ice + snow + graupel), shows that Thompson's FWC is dominated by snow, unlike M2005 and SAM1MOM. It is likely that, in Thompson, excessively efficient conversion of cloud ice to quickly falling snow causes the altitude of peak cloud fraction to be biased low.

Figure 4e shows longwave radiative cooling profiles for the simulations. Throughout most of their depth, cirrus clouds reduce radiative cooling by absorbing upwelling longwave radiation. M2005 has up to 0.5 K day^{-1} less radiative cooling than the other simulations between 8 and 13 km due to its comparably large cirrus coverage. Thompson and SAM1MOM, which have the smallest cirrus coverage, correspondingly have the strongest longwave cooling. These results are consistent with Hu et al. (2021).

Longwave CRE biases in the simulations can largely be explained by biases in the amount, the vertical structure, and the r_e of anvil cirrus, all of which can be estimated from spaceborne lidar and radar. These biases depend on the microphysics scheme; overall P3 best matches remote-sensing observations.

5 Simulated ice crystal populations lack observed variability

As a complementary test of the microphysics schemes, we compare simulated N_{ice} and FWC with in situ airborne observations from several tropical field studies, synthesized in the 'Microphysics Guide to Cirrus'(Krämer, Rolf, Spelten, Afchine, et al., 2020) (see Section 2, Text S3 and Figure S3), which have been coarsened to 0.04 Hz to match the horizontal grid spacing of the simulations. Observations are from heights above 10 km, and latitudes between 20°S and 20°N; model histograms are from high-cloud columns from all post-spin-up output times (day and night).

Figure 5 shows 2D histograms of FWC and N_{ice} for M2005, P3, Thompson and in situ observations. SAM1MOM is omitted because it does not predict or estimate N_{ice}. N_{ice} and FWC for M2005 and Thompson include cloud ice, graupel and snow. Vertical lines overlaid on the 2D histograms show limiters specified within the microphysics schemes. These limiters are designed to prevent algorithms within the schemes from producing physically implausible results; if the limiter is frequently active, this suggests problems with parameterization assumptions made within the scheme and/or biases in dynamics. Dotted lines show limiters on total cloud ice concentration and dashed lines show limiters on the concentration of ice particles produced through deposition nucleation, which is the dominant mode of nucleation within the temperature range investigated here. In Thompson, these two limiters are the same.

In M2005 and P3, most grid cells have values of N_{ice} that are very close to the smaller of these two limiters, which are 0.3 and 0.1 cm^{-3}, respectively. They have higher mean N_{ice} than the in situ observations and lack the observed variability in N_{ice} and dependence of N_{ice} on FWC.

Thompson has many grid cells with tiny FWC and N_{ice} and a subpopulation of grid cells dominated by snow (a large ratio of FWC to N_{ice}) as a result of efficiently converting most cloud ice to snow. Although P3 lacks the observed variability, its mean N_{ice} is closest to the observed mean.

6 Conclusions

Tropical longwave cloud radiative effects (CREs) simulated by a global storm-resolving model are highly sensitive to ice microphysics, even when nudging is used to largely remove microphysics-dynamics feedbacks. Average biases in longwave CRE vary over a 22 W m^{-2} range across four simulations which differ only in their microphysical schemes, due to variability in cirrus amount, thickness, cloud top height, and ice crystal number and size. This shows the need for further improvement of ice microphysics parameterizations, even in GSRMs, for which the convective forcing of cirrus clouds is much more realistically represented than in present-day GCMs.

Our study illustrates some key advantages of nudging, including isolating the sensitivity of simulated clouds to microphysics and precisely diagnosing radiative biases using coincident observations. However it also introduces a caveat, as nudging to a dataset that had lower vertical resolution than our model caused temperature biases around the cold point, which in turn caused tropical high cloud top heights to be biased low by nearly 2 km.

Simulations run with Thompson and SAM1MOM microphysics, which are partial double-moment and single-moment schemes, respectively, had weak longwave CREs. Thomp-

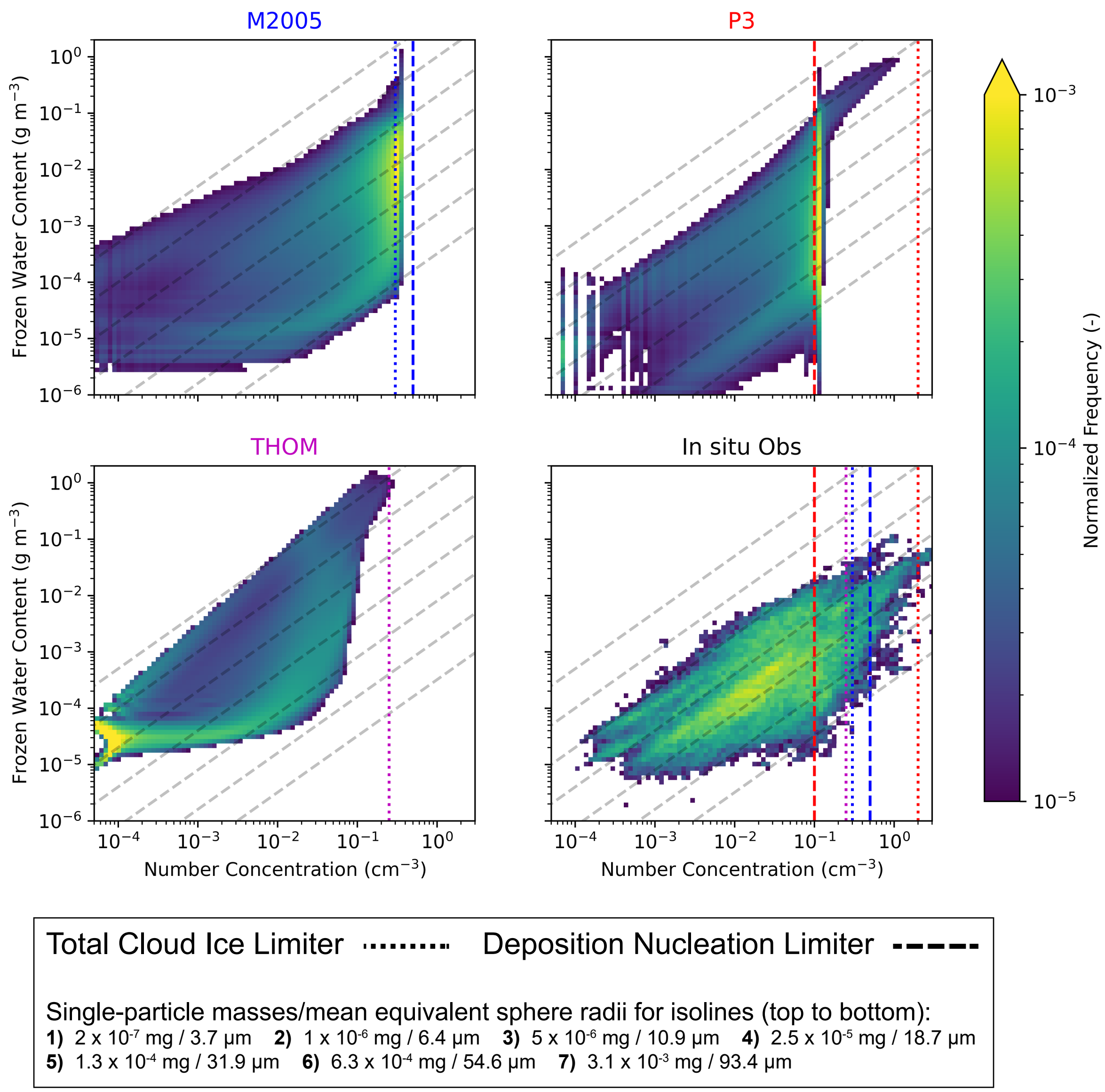


Figure 5. 2D histograms of FWC (y-axis, log-scale) and N_{ice} (x-axis, log-scale). Vertical dashed and dotted lines indicate limiters on total cloud ice number concentration and cloud ice particles that can be formed through deposition nucleation, respectively. Diagonal dashed lines show isolines; their corresponding single-particle masses and mean equivalent sphere radii are listed.

son rapidly converts cloud ice to larger snow particles, which fall quickly and reduce cirrus cloud cover, and decrease the optical depth of the remaining cirrus, even though the snow is radiatively active. SAM1MOM's small cirrus coverage may be related to the instantaneous sublimation of sedimenting cloud ice in subsaturated conditions, and to neglecting the radiative effects of snow.

The other two simulations, run with M2005 and P3 microphysics, which are both more complex double-moment schemes, had more cirrus and agreed better with climatological satellite retrieval products from DARDAR and 2C-ICE. P3's longwave CRE agrees best with coincident observations from the CERES satellite.

Simulated ice crystal number concentrations in M2005 and P3 ubiquitously hit arbitrary limiters within the microphysics schemes. As a result, typical ice crystal number concentrations lack the observed variability and dependence on frozen water content.

Ice crystal concentrations hitting limiters can result from too strong ice crystal sources, too weak ice crystal sinks and/or errors in the resolved dynamics. As M2005 and especially P3 are the most promising schemes, and are used in several existing GCMs and GSRMs, an important avenue for future work is detangling these factors to precisely diagnose the cause of too high ice crystal number concentrations. It is worth considering that the schemes used here have been developed mainly for the purpose of simulating midlatitude storm systems or, in the case of M2005, Arctic mixed-phase clouds (Morrison, Curry, Shupe, & Zuidema, 2005). Tropical high clouds likely have different dynamical and microphysical drivers. For example, convectively-generated gravity waves, which are only partly resolved by global storm-resolving models, are an important source of small-scale dynamic variability in the tropics (Atlas & Bretherton, 2023). Additionally, tropical high clouds exist at very cold temperatures and may be more influenced by homogeneous nucleation of aerosol, which is unrepresented in these schemes, and less influenced by heterogeneous nucleation.

7 Open Research

CERES(NASA/LARC/SD/ASDC, 2017), 2C-ICE R05 (`https://www.cloudsat.cira.colostate.edu/data-products/2c-ice`), DARDAR-CLOUD V2.1.0 and V3.10 (`http://www.icare.univ-lille1.fr`), and the Microphysics Guide to Cirrus (Krämer, Rolf, & Spelten, 2020) are publicly available online. Simulated model output cannot be made available due to the experimental nature of the simulations and the large storage space required.

Acknowledgments

We acknowledge funding from U.S. National Science Foundation (NSF) grants OISE-1743753 and AGS-1660604, high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by NSF, and we thank Blaž Gasparini and Maximilien Bolot for helpful discussions.

-

References

Atlas, R., & Bretherton, C. S. (2023). Aircraft observations of gravity wave activity and turbulence in the tropical tropopause layer: prevalence, influence on cirrus clouds, and comparison with global storm-resolving models [Journal Article]. *Atmos. Chem. Phys.*, *23*(7), 4009-4030. Retrieved from `https://acp.copernicus.org/articles/23/4009/2023/` (ACP) doi: 10.5194/acp-23-4009-2023

Atlas, R., Bretherton, C. S., Khairoutdinov, M. F., & Blossey, P. N. (2022). Hallett-Mossop rime splintering dims cumulus clouds over the Southern Ocean: New insight from nudged global storm-resolving simulations. *AGU Advances*, *3*(2), e2021AV000454. doi: https://doi.org/10.1029/2021AV000454

Bryan, G. H., Wyngaard, J. C., & Fritsch, J. M. (2003). Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review*, *131*(10), 2394-2416. doi: 10.1175/1520-0493(2003)131⟨2394:rrftso⟩2.0.co;2

Caldwell, P. M., Terai, C. R., Hillman, B., Keen, N. D., Bogenschutz, P., Lin, W., ... Zender, C. S. (2021). Convection-permitting simulations with the e3sm global atmosphere model [Journal Article]. *Journal of Advances in Modeling Earth Systems*, *13*(11), e2021MS002544. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002544` doi: https://doi.org/10.1029/2021MS002544

Cazenave, Q., Ceccaldi, M., Delanoë, J., Pelon, J., Groß, S., & Heymsfield, A. (2019). Evolution of DARDAR-CLOUD ice cloud retrievals: New parameters and impacts on the retrieved microphysical properties. *Atmos. Meas.*

Tech., *12*(5), 2819-2835. (AMT) doi: 10.5194/amt-12-2819-2019

Cooper, W. A. (1986). Ice initiation in natural clouds. *Meteorological Monographs*, *21*(43), 29-32. doi: 10.1175/0065-9401-21.43.29

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., ... Strand, W. G. (2020). The community earth system model version 2 (cesm2) [Journal Article]. *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001916` doi: 10.1029/2019ms001916

Delanoë, J., & Hogan, R. J. (2010). Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds. *Journal of Geophysical Research: Atmospheres*, *115*(D4). doi: https://doi.org/10.1029/2009JD012346

Deng, M., Mace, G. G., & Wang, Z. (2016). Anvil productivities of tropical deep convective clusters and their regional differences. *Journal of the Atmospheric Sciences*, *73*(9), 3467-3487. doi: 10.1175/jas-d-15-0239.1

Deng, M., Mace, G. G., Wang, Z., & Berry, E. (2015). CloudSat 2C-ICE product update with a new Ze parameterization in lidar-only region. *Journal of Geophysical Research: Atmospheres*, *120*(23), 12,198-12,208. doi: https://doi.org/10.1002/2015JD023600

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., ... Sun, M. (2013). Geostationary enhanced temporal interpolation for CERES flux products. *Journal of Atmospheric and Oceanic Technology*, *30*(6), 1072-1090. doi: 10.1175/jtech-d-12-00136.1

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., ... Zhu, Q. (2019). The doe e3sm coupled model version 1: Overview and evaluation at standard resolution [Journal Article]. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089-2129. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001603` doi: https://doi.org/10.1029/2018MS001603

Hartmann, D. L., Moy, L. A., & Fu, Q. (2001). Tropical convection and the energy balance at the top of the atmosphere. *Journal of Climate*, *14*(24), 4495-4511. doi: 10.1175/1520-0442(2001)014⟨4495:tcateb⟩2.0.co;2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,

... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: 10.1002/qj.3803

Hong, Y., Liu, G., & Li, J.-L. F. (2016). Assessing the radiative effects of global ice clouds based on CloudSat and CALIPSO measurements. *Journal of Climate*, *29*(21), 7651-7674. doi: 10.1175/jcli-d-15-0799.1

Hu, Z., Lamraoui, F., & Kuang, Z. (2021). Influence of upper-troposphere stratification and cloud–radiation interaction on convective overshoots in the tropical tropopause layer. *Journal of the Atmospheric Sciences*, *78*(8), 2493-2509. doi: 10.1175/jas-d-20-0241.1

Jensen, E. J., Pfister, L., Jordan, D. E., Bui, T. V., Ueyama, R., Singh, H. B., ... Pfeilsticker, K. (2017). The NASA Airborne Tropical Tropopause Experiment: High-altitude aircraft measurements in the tropical western Pacific. *Bulletin of the American Meteorological Society*, *98*(1), 129-+. doi: 10.1175/bams-d-14-00263.1

Khairoutdinov, M. F., Blossey, P. N., & Bretherton, C. S. (2022). Global system for atmospheric modeling: Model description and preliminary results. *Earth and Space Science Open Archive*, 39. doi: doi:10.1002/essoar.10509977.1

Khairoutdinov, M. F., & Randall, D. A. (2003). Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *Journal of the Atmospheric Sciences*, *60*(4), 607-625. doi: 10.1175/1520-0469(2003)060⟨0607:crmota⟩2.0.co;2

Krämer, M., Rolf, C., & Spelten, N. (2020). *The Cirrus Guide II In-situ Aircraft Data Set.* doi: 10.34730/266ca2a41f4946ff97d874bfa458254c

Krämer, M., Rolf, C., Spelten, N., Afchine, A., Fahey, D., Jensen, E., ... Sourdeval, O. (2020). A microphysics guide to cirrus – Part 2: Climatologies of clouds and humidity from observations. *Atmos. Chem. Phys.*, *20*(21), 12569-12608. (ACP) doi: 10.5194/acp-20-12569-2020

Lohmann, U., Spichtinger, P., Jess, S., Peter, T., & Smit, H. (2008). Cirrus cloud formation and ice supersaturated regions in a global climate model. *Environmental Research Letters*, *3*(4). doi: 10.1088/1748-9326/3/4/045022

Morrison, H., Curry, J. A., & Khvorostyanov, V. I. (2005). A new double-moment microphysics parameterization for application in cloud and climate models. Part I: description. *Journal of the Atmospheric Sciences*, *62*(6), 1665-1677.

doi: 10.1175/jas3446.1

Morrison, H., Curry, J. A., Shupe, M. D., & Zuidema, P. (2005). A new double-moment microphysics parameterization for application in cloud and climate models. part ii: Single-column modeling of arctic clouds [Journal Article]. *Journal of the Atmospheric Sciences*, *62*(6), 1678-1693. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/62/6/jas3447.1.xml` doi: https://doi.org/10.1175/JAS3447.1

Morrison, H., & Milbrandt, J. A. (2015). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: scheme description and idealized tests. *Journal of the Atmospheric Sciences*, *72*(1), 287-311. doi: 10.1175/jas-d-14-0065.1

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review*, *137*(3), 991-1007. doi: 10.1175/2008mwr2556.1

NASA/LARC/SD/ASDC. (2017). *CERES and GEO-enhanced TOA, within-atmosphere and surface fluxes, clouds and aerosols 1-hourly Terra Edition4A.* NASA Langley Atmospheric Science Data Center DAAC. doi: 10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A

Pan, L. L., Atlas, E. L., Salawitch, R. J., Honomichl, S. B., Bresch, J. F., Randel, W. J., … Wolfe, G. (2017). The Convective Transport of Active Species in the Tropics (CONTRAST) Experiment. *Bulletin of the American Meteorological Society*, *98*(1), 106-128. doi: 10.1175/bams-d-14-00272.1

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., … Zelinka, M. D. (2020). An assessment of earth's climate sensitivity using multiple lines of evidence [Journal Article]. *Reviews of Geophysics*, *58*(4), e2019RG000678. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019RG000678` doi: https://doi.org/10.1029/2019RG000678

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications [Journal Article]. *Journal of Computational Physics*, *227*(7), 3465-3485. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0021999107000459` doi:

https://doi.org/10.1016/j.jcp.2007.01.037

Skamarock, W. C., Klemp, J. B., Duda, M. G., Fowler, L. D., Park, S.-H., & Ringler, T. D. (2012). A multiscale nonhydrostatic atmospheric model using centroidal voronoi tesselations and c-grid staggering [Journal Article]. *Monthly Weather Review*, *140*(9), 3090-3105. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/140/9/mwr-d-11-00215.1.xml` doi: https://doi.org/10.1175/MWR-D-11-00215.1

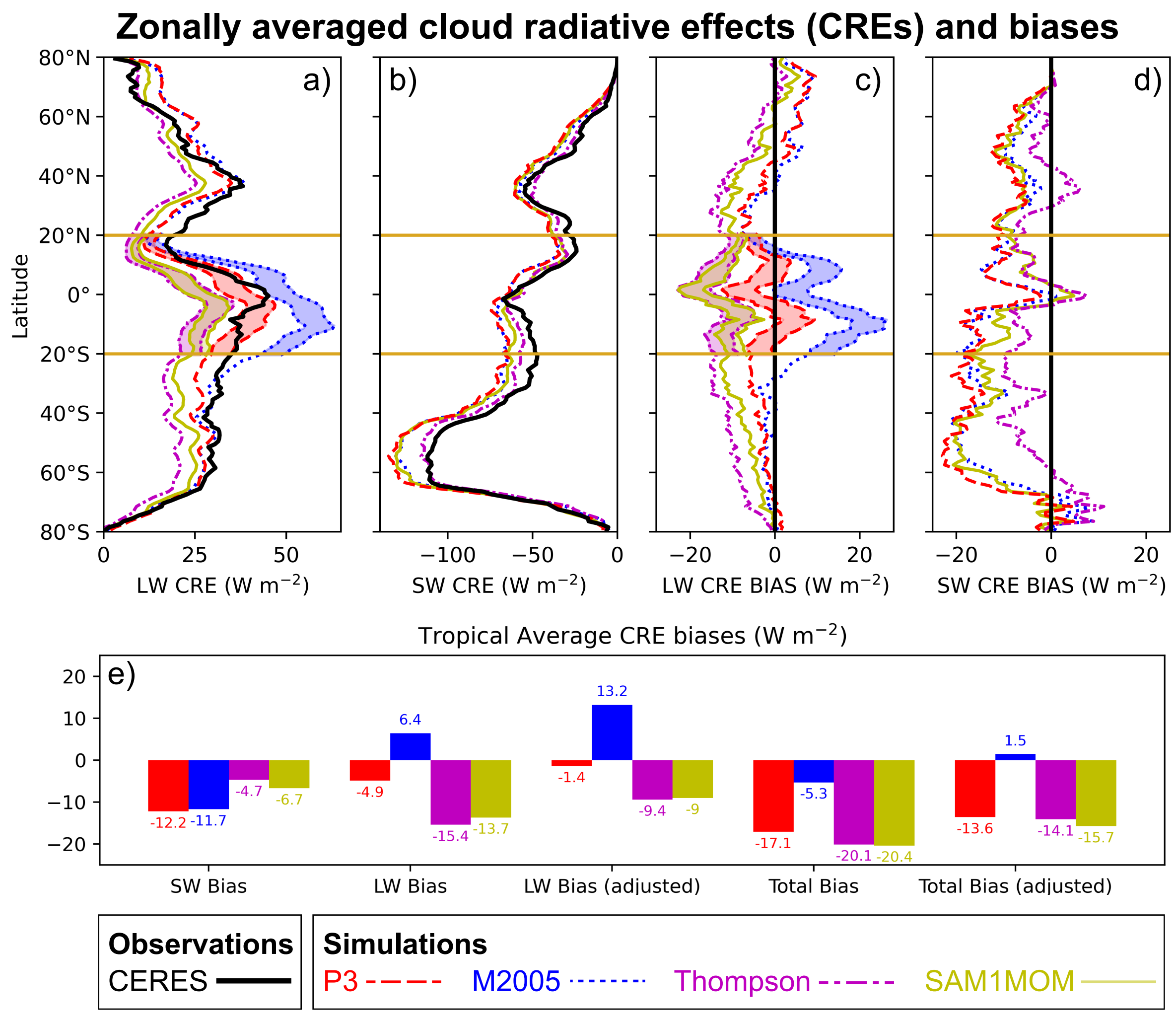
Stanford, M. W., Varble, A., Zipser, E., Strapp, J. W., Leroy, D., Schwarzenboeck, A., ... Protat, A. (2017). A ubiquitous ice size bias in simulations of tropical deep convection. *Atmos. Chem. Phys.*, *17*(15), 9599-9621. (ACP) doi: 10.5194/acp-17-9599-2017

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*(12), 5095-5115. doi: 10.1175/2008mwr2387.1

Toon, O. B., Starr, D. O., Jensen, E. J., Newman, P. A., Platnick, S., Schoeberl, M. R., ... Pickering, K. E. (2010). Planning, implementation, and first results of the Tropical Composition, Cloud and Climate Coupling Experiment (TC4). *Journal of Geophysical Research: Atmospheres*, *115*(D10). doi: https://doi.org/10.1029/2009JD013073

Wendisch, M., Pöschl, U., Andreae, M. O., Machado, L. A. T., Albrecht, R., Schlager, H., ... Zöger, M. (2016). ACRIDICON–CHUVA Campaign: Studying tropical deep convective clouds and precipitation over Amazonia using the new German research aircraft HALO. *Bulletin of the American Meteorological Society*, *97*(10), 1885-1908. doi: 10.1175/bams-d-14-00255.1

Figure 1.



Zonally averaged cloud radiative effects (CREs) and biases
a)
b)
c)
d)
Latitude
80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
80°S
LW CRE (W m^{-2})
SW CRE (W m^{-2})
LW CRE BIAS (W m^{-2})
SW CRE BIAS (W m^{-2})
Tropical Average CRE biases (W m^{-2})
e)
SW Bias
LW Bias
LW Bias (adjusted)
Total Bias
Total Bias (adjusted)
-12.2
-11.7
-4.7
-6.7
-4.9
6.4
-15.4
-13.7
-1.4
13.2
-9.4
-9
-17.1
-5.3
-20.1
-20.4
-13.6
1.5
-14.1
-15.7
Observations
CERES
Simulations
P3
M2005
Thompson
SAM1MOM

Figure 2.

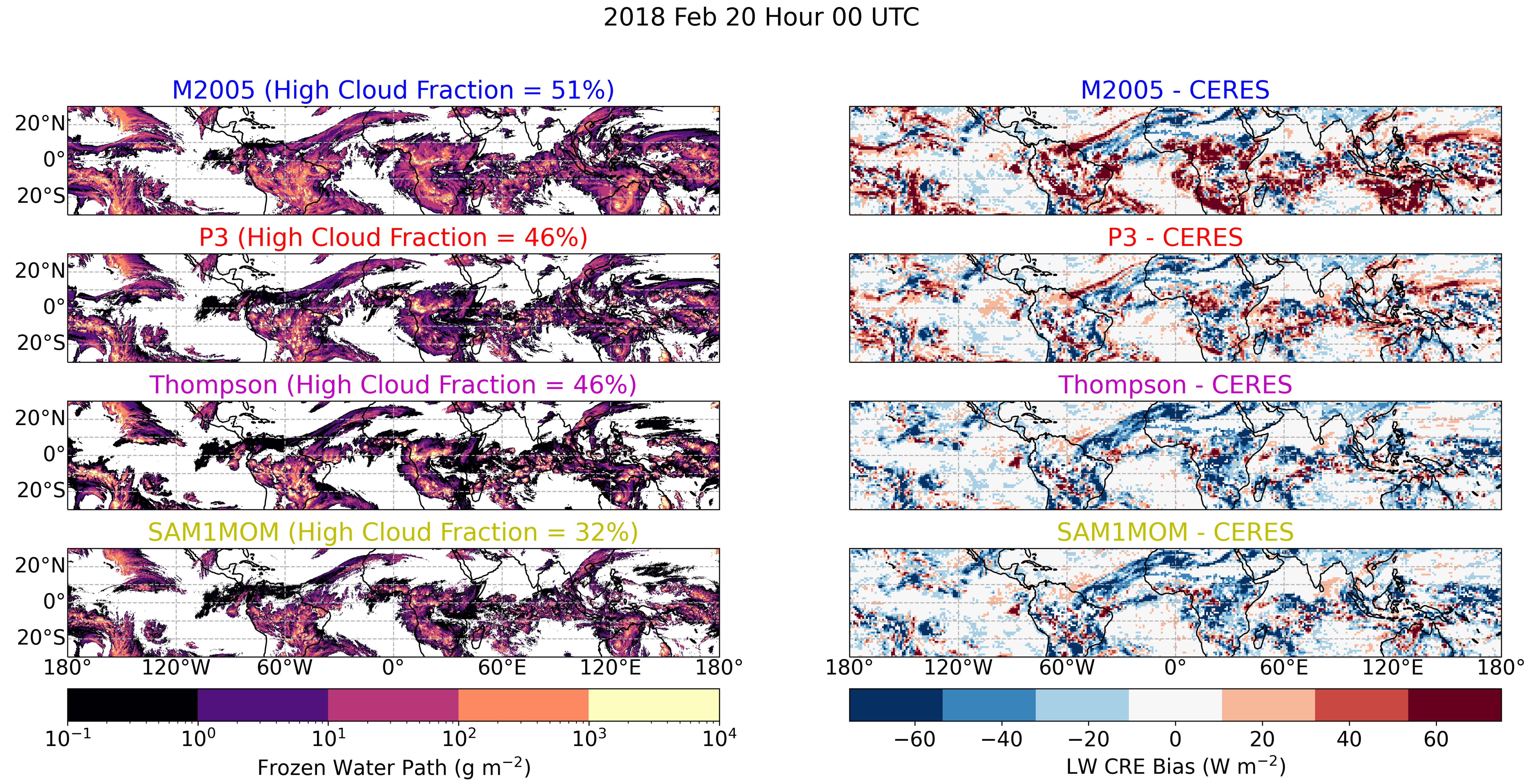
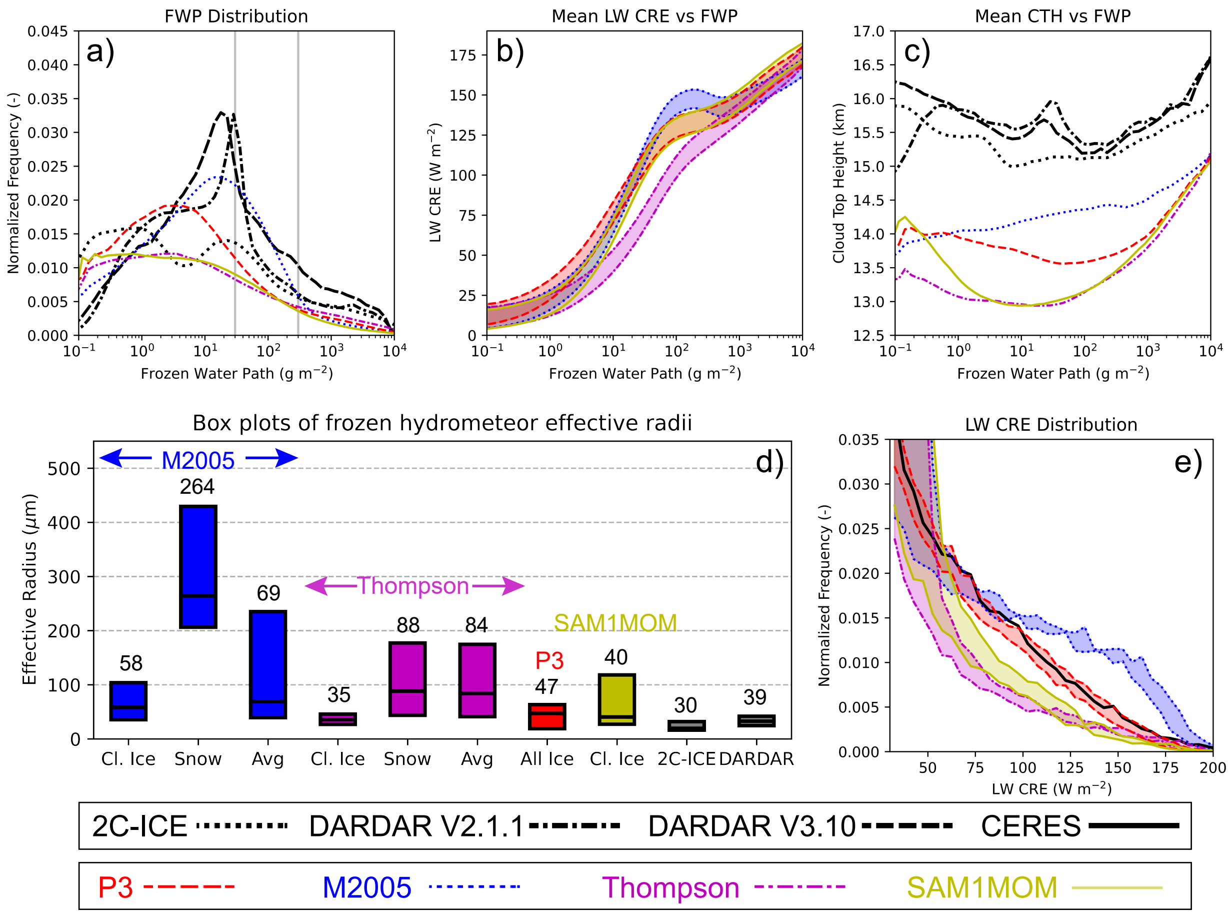
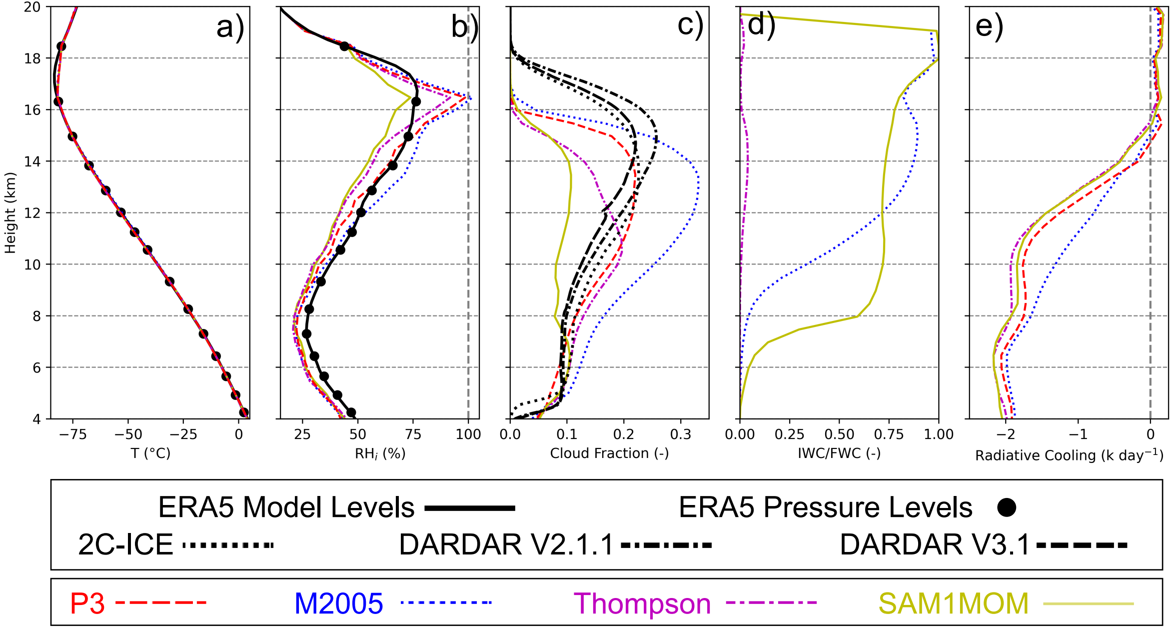

2018 Feb 20 Hour 00 UTC
M2005 (High Cloud Fraction = 51%)
M2005 - CERES
P3 (High Cloud Fraction = 46%)
P3 - CERES
Thompson (High Cloud Fraction = 46%)
Thompson - CERES
SAM1MOM (High Cloud Fraction = 32%)
SAM1MOM - CERES
20°N
0°
20°S
180°
120°W
60°W
0°
60°E
120°E
180°
10^−1
10^0
10^1
10^2
10^3
10^4
Frozen Water Path (g m^−2)
−60
−40
−20
0
20
40
60
LW CRE Bias (W m^−2)

Figure 3.



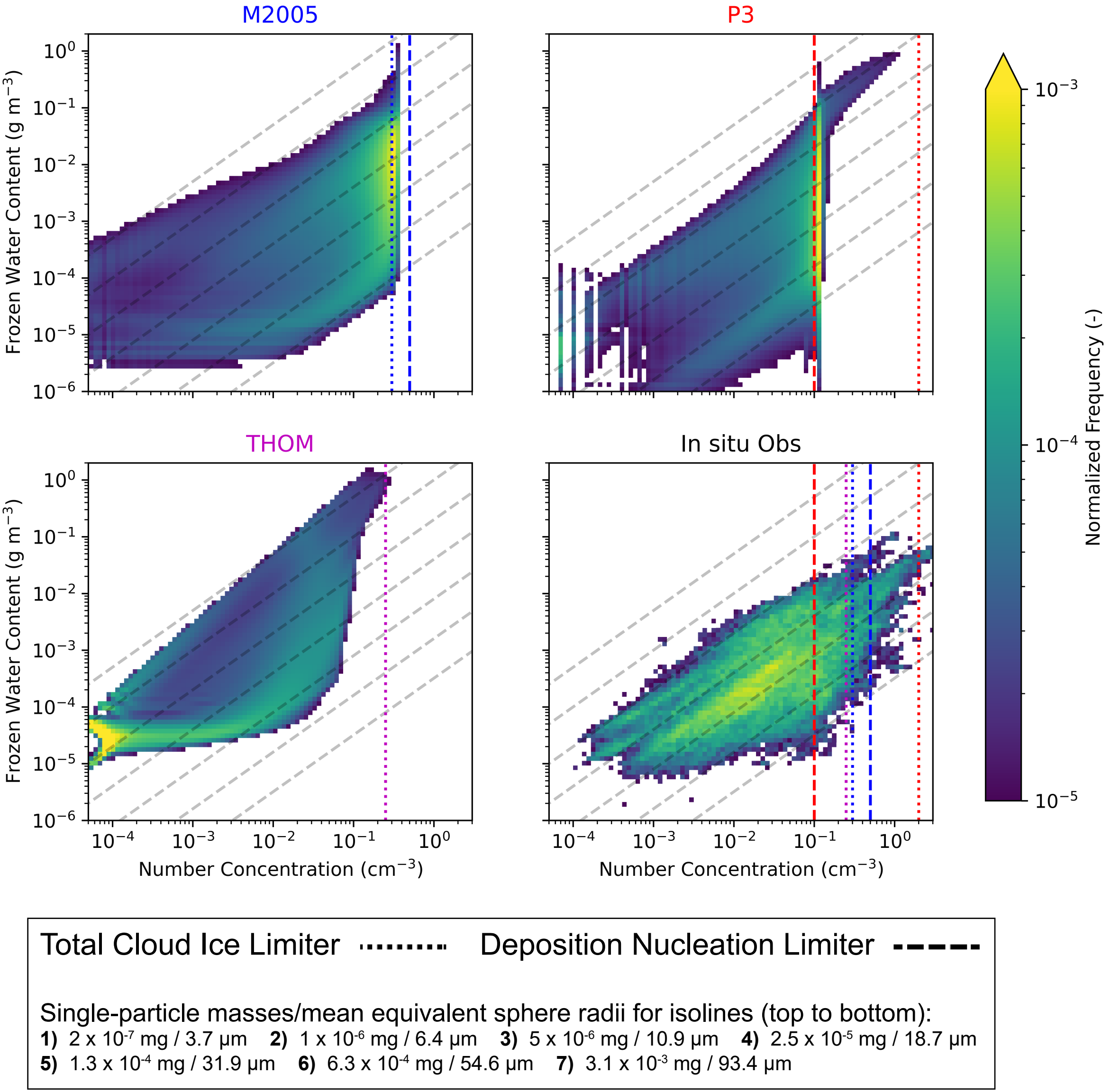
FWP Distribution
a)
Normalized Frequency (-)
Frozen Water Path (g m^{-2})
Mean LW CRE vs FWP
b)
LW CRE (W m^{-2})
Frozen Water Path (g m^{-2})
Mean CTH vs FWP
c)
Cloud Top Height (km)
Frozen Water Path (g m^{-2})
Box plots of frozen hydrometeor effective radii
d)
M2005
Thompson
SAM1MOM
P3
Effective Radius (μm)
58
264
69
35
88
84
47
40
30
39
Cl. Ice
Snow
Avg
Cl. Ice
Snow
Avg
All Ice
Cl. Ice
2C-ICE
DARDAR
LW CRE Distribution
e)
Normalized Frequency (-)
LW CRE (W m^{-2})
2C-ICE
DARDAR V2.1.1
DARDAR V3.10
CERES
P3
M2005
Thompson
SAM1MOM

Figure 4.



a)
b)
c)
d)
e)
Height (km)
T (°C)
RH$_i$ (%)
Cloud Fraction (-)
IWC/FWC (-)
Radiative Cooling (k day^{-1})
ERA5 Model Levels
ERA5 Pressure Levels
2C-ICE
DARDAR V2.1.1
DARDAR V3.1
P3
M2005
Thompson
SAM1MOM

Figure 5.



M2005
P3
THOM
In situ Obs
Frozen Water Content (g m^{-3})
Number Concentration (cm^{-3})
Normalized Frequency (-)
Total Cloud Ice Limiter
Deposition Nucleation Limiter
Single-particle masses/mean equivalent sphere radii for isolines (top to bottom):
1) 2 x 10^{-7} mg / 3.7 μm **2)** 1 x 10^{-6} mg / 6.4 μm **3)** 5 x 10^{-6} mg / 10.9 μm **4)** 2.5 x 10^{-5} mg / 18.7 μm
5) 1.3 x 10^{-4} mg / 31.9 μm **6)** 6.3 x 10^{-4} mg / 54.6 μm **7)** 3.1 x 10^{-3} mg / 93.4 μm

Supporting Information for "Tropical anvil cirrus are highly sensitive to ice microphysics within a nudged global storm-resolving model"

R.L. Atlas[1], C.S. Bretherton[2], A.B. Sokol[1], P.N. Blossey[1], M.F. Khairoutdinov[3]

[1]Dept. of Atmospheric Sciences, University of Washington, Seattle, WA

[2]Allen Institute for Artificial Intelligence, Seattle, WA

[3]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY

[1]3920 Okanogan Ln, Seattle, WA 98195

[2]2157 N Northlake Way #110, Seattle, WA 98103

[3]Endeavour, 145, Stony Brook, NY 11790

Contents of this file

1. Text S1 to S4
2. Figures S1 to S5

Additional Supporting Information (File uploaded separately)

1. Caption for Movie S1

Text S1: Description of ice microphysics in the four different microphysics schemes

In SAM1MOM (Khairoutdinov & Randall, 2003), two prognostic variables represent all water species: (1) total water mass mixing ratio, which combines water vapor and non-precipitating hydrometeors and (2) the precipitating hydrometeor mass mixing ratio. Both non-precipitating (cloud liquid and cloud ice) and precipitating (rain, snow and graupel) hydrometeors are partitioned between liquid and ice phases based on temperature, and ice phase precipitating hydrometeor mass is further partitioned between snow and graupel based on temperature. Only cloud ice is radiatively active. SAM1MOM partitions total water into water vapor and cloud condensate using saturation adjustment at all temperatures, including for cloud ice. This means that cloud ice condenses and sublimates instantaneously at ice saturation. Rain, snow and graupel number are prescribed as functions of rain, snow and graupel mass, respectively, but cloud ice number does not exist in the scheme.

M2005 (SAM version 3.5), which is the microphysics scheme originally described in Morrison, Curry, and Khvorostyanov (2005) but with a rimed ice species added as in Morrison, Thompson, and Tatarskii (2009), predicts number and mass for three frozen hydrometeor classes (cloud ice, snow and graupel). Cloud ice and snow are both radiatively active. Thompson (Thompson et al., 2008) (based on WRF version 3.5.1) predicts mass for three frozen hydrometeor classes (cloud ice, snow and graupel) and number for cloud ice only. Snow number is prescribed as a function of snow mass and temperature following Field et al. (2005). Cloud ice and snow are both radiatively active. P3 (Morrison & Milbrandt, 2015) (SAM version and 3.5) is run with one radiatively active ice class, for which it predicts mass, number, rime volume and rime mass.

M2005, Thompson and P3 heterogeneously nucleate ice through deposition and immersion freezing. M2005 also includes contact nucleation. At the temperatures and heights examined here, deposition nucleation dominates heterogeneous nucleation.

In M2005, deposition nucleation occurs when either ice supersaturation exceeds 8% or the air is saturated with respect to liquid and colder than $-8°$C. In Thompson, it occurs when either ice supersaturation exceeds 25% or air is saturated with respect to liquid and colder than $-12°$C. In P3, it occurs when the temperature is below $-15°$C and ice supersaturation exceeds 5%. All three schemes use the Cooper curve (Cooper, 1986) to specify the concentration of ice nucleating particles for deposition nucleation and have limiters which specify a maximum concentration of ice particles that can be formed by deposition nucleation. The limiters in P3, Thompson, and M2005 are .1, .25 and .5 cm^{-3}, respectively.

All three microphysics schemes also support homogeneous freezing of droplets and raindrops when the air temperature is $< -40°$C but do not support homogeneous freezing of aerosol. Limiters act to restrict the total concentrations of cloud ice particles to be no larger than 2, .25 and .3 cm^{-3} in P3, Thompson and M2005, respectively.

Text S2: Processing of DARDAR and 2C-ICE

DARDAR and 2C-ICE both retrieve frozen water content (FWC) from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) lidar attenuated backscatter and CloudSat radar reflectivity. A major difference between the two retrievals is that 2C-ICE parameterizes radar reflectivity for grid cells where the cloud is too thin to be detected by the CloudSat radar (Deng et al., 2015). Here, we compare retrievals of frozen water content (FWC) and frozen water path (FWP) between DARDAR V3.10 and 2C-ICE to examine the impact of

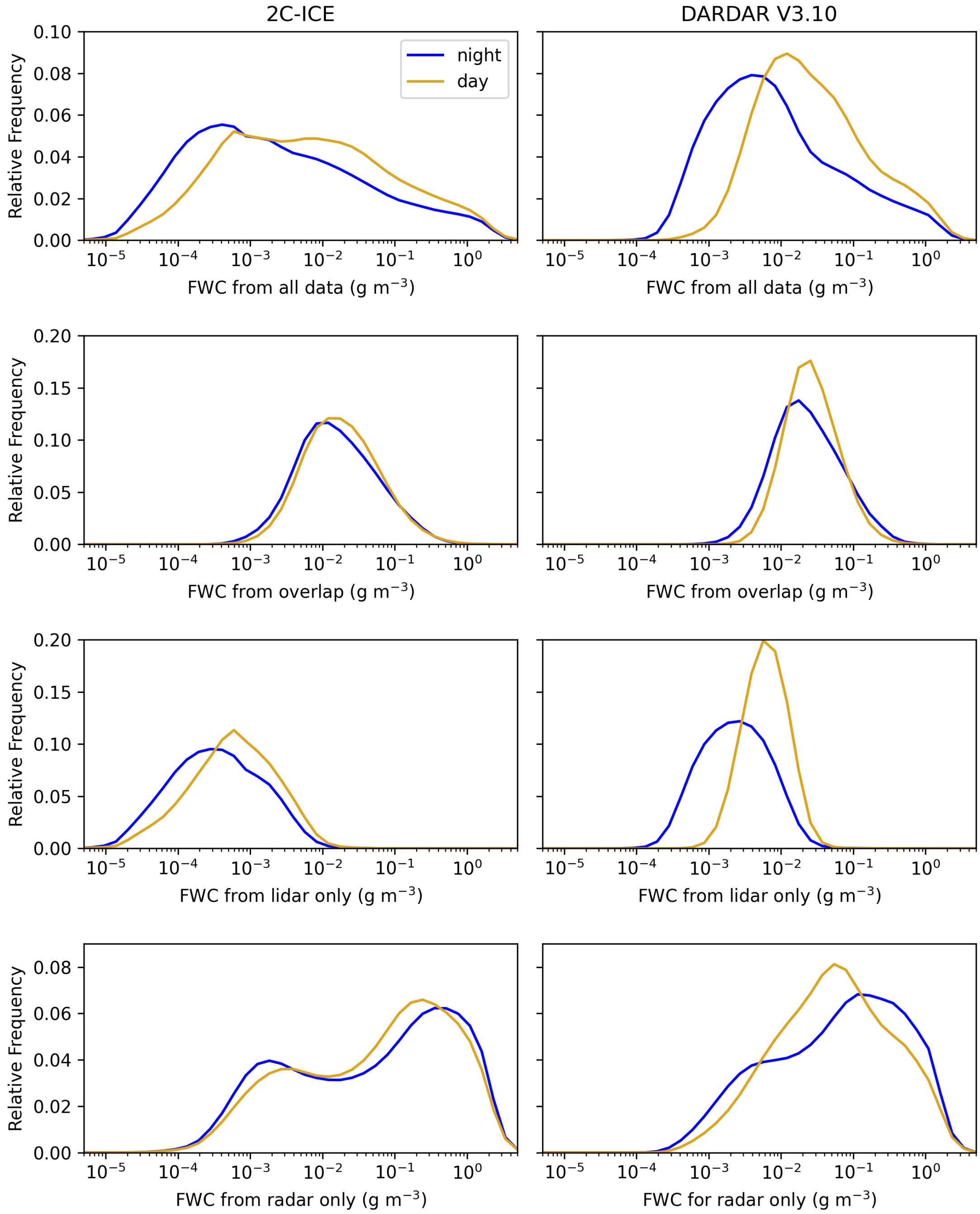


Figure S1. Distributions of FWC from nighttime and daytime measurements separately for (top to bottom row) all data, regions sensed by both the radar and lidar, lidar only regions, and radar only regions, for 2C-ICE (left) and DARDAR (right).

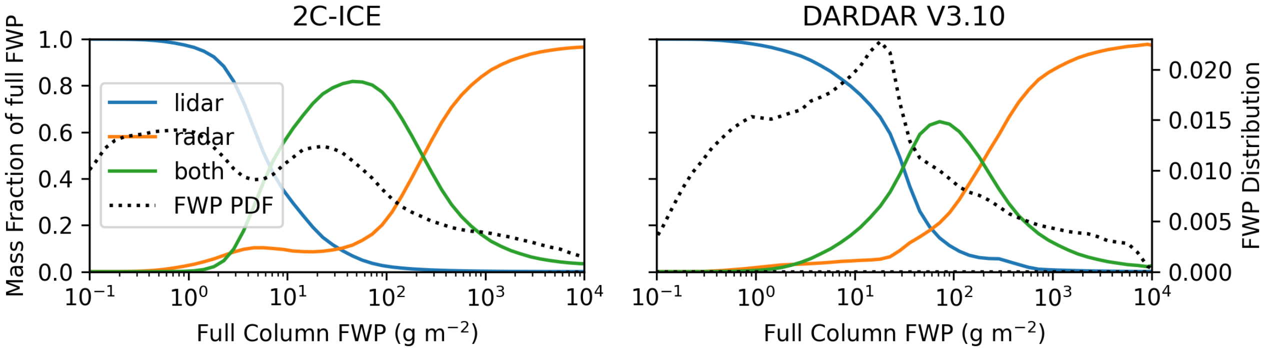


Figure S2. Blue, orange and green lines show the average mass fraction of a column that is sensed by the lidar only, radar only, and both instruments, respectively, as a function of column FWP.

that difference. Because the two versions of DARDAR are more similar to each other than they are to the 2C-ICE, we only examine the newer version of DARDAR here.

Figure S1 shows distributions of FWC from the two satellite retrievals broken up into daytime and nighttime measurements, and, in the bottom three rows, according to which instruments the retrieval is coming from (lidar only, radar only or both). In general, retrieved FWCs are smaller in 2C-ICE than in DARDAR. Most of this difference comes from lidar-only regions, where 2C-ICE returns FWCs that are about one order of magnitude smaller on average than those retrieved by DARDAR. 2C-ICE also has a more bimodal distribution than DARDAR for radar-only regions. The two retrievals agree best for regions with both instruments.

DARDAR's retrievals show a greater diurnal dependence, particular in the lidar-only regions, due to the fact that the lidar is more sensitive at night. Because DARDAR has greater sensitivity at night, we restrict our comparisons between the simulations and satellite retrievals to nighttime measurements. Additionally, because DARDAR cannot detect FWCs $< 10^{-4}$ g m^{-3} at night, we filter FWCs smaller than that out of both the simulated output and the satellite retrievals before computing FWP.

Given that the retrievals diverge most from each other in lidar only regions, we examine the mass fraction that comes from lidar-only regions, radar-only regions and regions with both instruments as a function of FWP in Figure S2 (left y-axis). Distributions of FWP are overlaid (right y-axis). For FWPs > 30 g m^{-2}, most of the FWP comes from regions with both instruments or with radar only. Accordingly, the two retrievals agree well within this range. For FWPs < 30 g m^{-2}, the satellite retrievals are very different from each other and do not provide as tight a constraint on the simulations.

Text S3: Processing of the 'Microphysics Guide to Cirrus'

The 'Microphysics Guide to Cirrus' (Krämer, Rolf, Spelten, Afchine, et al., 2020; Krämer, Rolf, & Spelten, 2020) includes quality controlled microphysics and thermodynamics observations from 24 field campaigns. Five of those campaigns measured FWC and ice crystal number concentration (N_{ice}) at latitudes between 20°S and 20°N and altitudes > 10 km, including Airborne Tropical TRopopause EXperiment (Jensen et al., 2017, ATTREX), Convective Transport of Active Species in the Tropics EXperiment (Pan et al., 2017, CONTRAST), Aerosol, Cloud, Precipitation, and Radiation Interactions and Dynamics of Convective Cloud Systems (Wendisch et al., 2016, ACRIDICON), Tropical Composition, Cloud and Climate Coupling Experiment (Toon et al., 2010, TC4), and Pacific Oxidants, Sulfur, Ice, Dehydration, and cONvection (POSIDON). Figure S3 shows the flight tracks from all five campaigns, and lists the instruments used to measure or compute FWC and N_{ice}.

All data in the Microphysics Guide have a resolution of 1 Hz. Air speeds in the upper troposphere are typically 200 m s^{-1}, so we coarsened the data to .04 Hz (or 25 seconds) so that each data point would correspond to an approximately 5 km horizontal distance, and better match

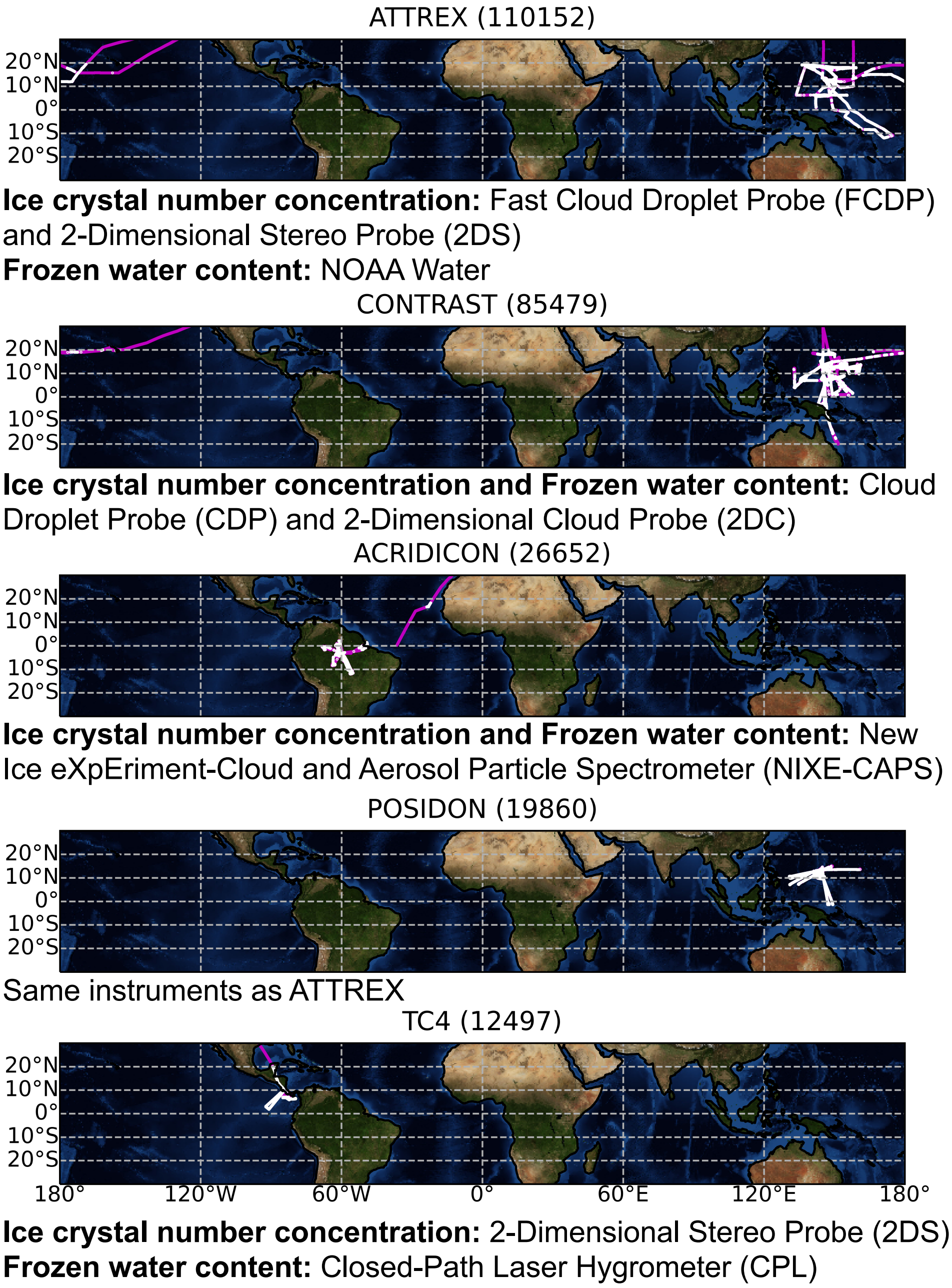


Figure S3. Campaign flight tracks in magenta with white overlay indicating in-cloud data above 10 km and within 20°N and 20°S. Map titles include the campaign name and the number of .04 Hz data points used in parentheses. Below each map, instruments used to measure or compute FWC and N_{ice} are listed.

the spatial scale of the simulated output. The numbers next to the flight campaign names in Figure S3 are the number of in-cloud, 0.04 Hz data points that match the latitude and altitude criteria.

Text S4: Cloud top height bias correction

As discussed in Section 4 of the main text, and shown in Figures 3 and 4, cloud top heights (CTHs) are biased low in all simulations due to temperature biases around the cold point inherited, through nudging, from the ERA5 reanalysis data on pressure levels. We correct for this bias by computing an average cloud top temperature (CTT) bias from the the average CTH bias and the average CTT lapse rate in the simulations. The left column of Figure S4 shows distributions of CTH for simulations and observations for land and ocean separately. The CTH bias in present over both land and ocean but is slightly smaller over land. The average CTH bias is 1.85 km. The right column of the Figure S4 shows CTT as a function of CTH. The average CTT lapse rate is 7.6 K km^{-1}. Multiplying the average CTH bias by the average CTT lapse rate, we get that the average CTT bias in the simulations is 14 K. We adjust the longwave CREs for clouds with CTH > 10 km, to account for the CTH bias, using the following equation:

$$\text{Adjusted longwave CRE} = \text{longwave CRE} + \epsilon \times k_B \times [\text{CTT}^4 - (\text{CTT} - 14)^4]$$

ϵ is the cirrus emissivity, which we estimate to be 0.5.

Figures S5 is a version of Figure 2 using the adjusted longwave CREs. The low CTH bias artificially reduces the longwave CREs in the simulations, so accounting for it increases them. This improves the longwave CRE biases in P3, Thompson and SAM1MOM, but makes them worse in M2005. SAM1MOM and Thompson still have longwave CREs that are substantially too low after the adjustment.

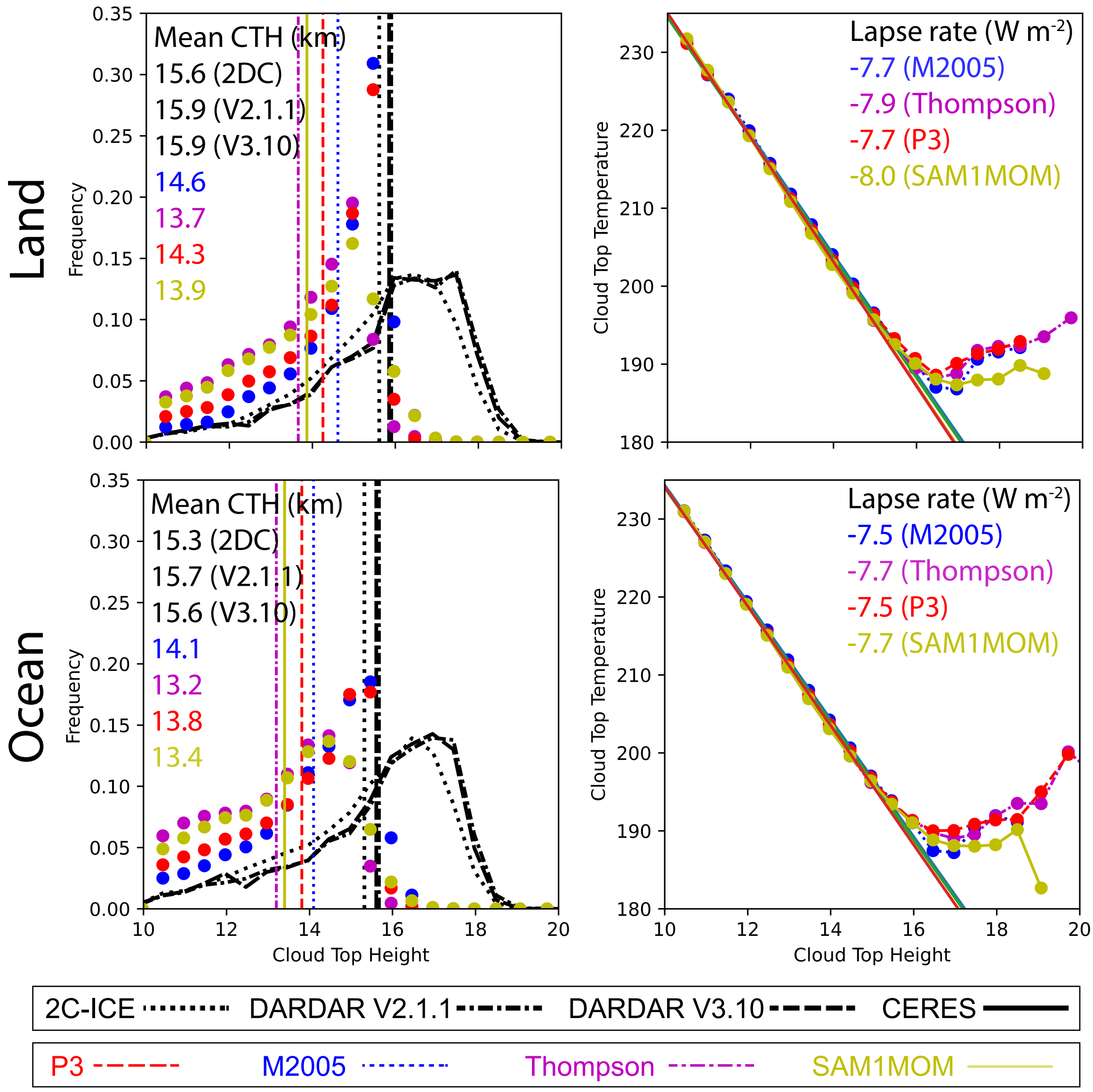


Figure S4. Left column shows normalized distributions of cloud top height (CTH) for the observations and simulations for land (top) and ocean (bottom) separately. Mean cloud top heights are indicated by the vertical lines and printed on the plot. Right column shows mean cloud top temperature (CTT) as a function of height. Lines are fitted for CTHs between 10 and 14 km to estimate CTT lapse rates, which are printed on the plots.

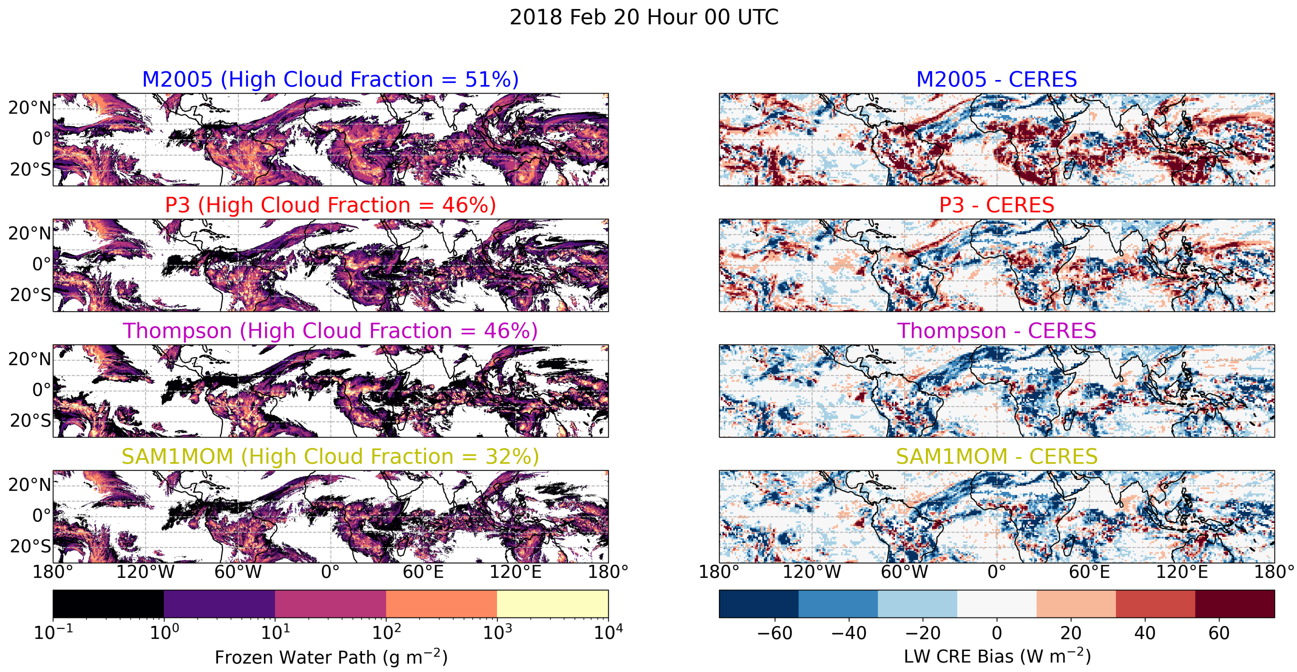


Figure S5. Same as Figure 2 but using the adjusted longwave CRE

Movie S1. For each hour of output from days 2-5 of the simulations, we show **left:** Snapshots of simulated frozen water path (FWP, including cloud ice, snow and graupel) for columns with cloud top height (CTH) $>$ 10 km on the simulations' native grid and **right:** Coincident snapshots of longwave CRE bias compared to CERES on a coarsened 1° x 1° grid. At high zenith angles, CERES sometimes mistakes land for cloud, causing a positive (blue) bias over the land. This is especially evident over Africa.