How well do large-eddy simulations and global climate models represent observed boundary layer structures and low clouds over the summertime Southern Ocean?

Rachel Atlas[1], Christopher S. Bretherton[1], Peter N. Blossey[1], Andrew Gettelman[2], Charles Bardeen[2], Pu Lin[3], and Yi Ming[4]

[1]University of Washington
[2]National Center for Atmospheric Research (UCAR)
[3]Princeton University
[4]Geophysical Fluid Dynamics Laboratory

November 24, 2022

Abstract

Climate models struggle to accurately represent the highly reflective boundary layer clouds overlying the remote and stormy Southern Ocean. We use in-situ aircraft observations from the Southern Ocean Clouds, Radiation and Aerosol Transport Experimental Study (SOCRATES) to evaluate Southern Ocean clouds in a cloud-resolving large-eddy simulation (LES) and two coarse resolution global atmospheric models, the CESM Community Atmosphere Model (CAM6) and the GFDL global atmosphere model (AM4), run in a nudged hindcast framework. We develop six case studies from SOCRATES data which span the range of observed cloud and boundary layer properties. For each case, the LES is run once forced purely using reanalysis data ('ERA5-based') and once strongly nudged to an aircraft profile ('Obs-based'). The ERA5-based LES can be compared with the global models, which are also nudged to reanalysis data, and is better for simulating cumulus. The Obs-based LES closely matches an observed cloud profile and is useful for microphysical comparisons and sensitivity tests, and simulating multi-layer stratiform clouds. We use two-moment Morrison microphysics in the LES and find that it simulates too few frozen particles in clouds occurring within the Hallett-Mossop temperature range. We modify the Hallett-Mossop parameterization so that it activates within boundary layer clouds and we achieve better agreement between observed and simulated microphysics. The nudged GCMs achieve reasonable supercooled liquid water dominated clouds in most cases but struggle to represent multi-layer stratiform clouds and to maintain liquid water in cumulus clouds. CAM6 has low droplet concentrations in all cases and underestimates stratiform cloud-driven turbulence.

How well do large-eddy simulations and global climate models represent observed boundary layer structures and low clouds over the summertime Southern Ocean?

R.L. Atlas[1], C.S. Bretherton[1,2], P.N. Blossey[1], A. Gettelman[3], C. Bardeen[3], Pu Lin[4], Yi Ming[4]

[1]Dept. of Atmospheric Sciences, University of Washington, Seattle, WA
[2]Vulcan Climate Modeling, Seattle, WA
[3]National Center for Atmospheric Research (NCAR), Boulder, CO
[4]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey
[5]Geophysical Fluid Dynamics Laboratory (GFDL)

Key Points:

- SAM LES represents diverse Southern Ocean boundary layer and cloud structures very well
- CAM6 and AM4 maintain supercooled liquid water in stratiform clouds but excessively glaciate cumuli
- CAM6 is deficient in stratiform cloud-driven turbulence and cloud droplets

Corresponding author: Rachel Atlas, `ratlas@uw.edu`

Abstract

Climate models struggle to accurately represent the highly reflective boundary layer clouds overlying the remote and stormy Southern Ocean. We use in-situ aircraft observations from the Southern Ocean Clouds, Radiation and Aerosol Transport Experimental Study (SOCRATES) to evaluate Southern Ocean clouds in a cloud-resolving large-eddy simulation (LES) and two coarse resolution global atmospheric models, the CESM Community Atmosphere Model (CAM6) and the GFDL global atmosphere model (AM4), run in a nudged hindcast framework. We develop six case studies from SOCRATES data which span the range of observed cloud and boundary layer properties. For each case, the LES is run once forced purely using reanalysis data ('ERA5-based') and once strongly nudged to an aircraft profile ('Obs-based'). The ERA5-based LES can be compared with the global models, which are also nudged to reanalysis data, and is better for simulating cumulus. The Obs-based LES closely matches an observed cloud profile and is useful for microphysical comparisons and sensitivity tests, and simulating multi-layer stratiform clouds. We use two-moment Morrison microphysics in the LES and find that it simulates too few frozen particles in clouds occurring within the Hallett-Mossop temperature range. We modify the Hallett-Mossop parameterization so that it activates within boundary layer clouds and we achieve better agreement between observed and simulated microphysics. The nudged GCMs achieve reasonable supercooled liquid water dominated clouds in most cases but struggle to represent multi-layer stratiform clouds and to maintain liquid water in cumulus clouds. CAM6 has low droplet concentrations in all cases and underestimates stratiform cloud-driven turbulence.

Plain Language Summary

The Southern Ocean, the wide band of water North of Antarctica, is the stormiest place on Earth. Weather systems constantly whirl the atmosphere and blanket the ocean in clouds. Low-lying clouds reflect sunlight back to space and cool the Earth. Here, we investigate how well the computer models that we use to understand the climate and to forecast future climates can simulate these clouds.

We use recent aircraft measurements from the Southern Ocean Clouds, Radiation, Aerosol Transport Experimental Study (SOCRATES) to evaluate two leading U.S. global climate models, the GFDL global atmosphere model (AM4) and the CESM Community Atmosphere Model (CAM6). We additionally run detailed simulations of Southern Ocean

clouds over a small area, to understand which physical processes are relevant to cloud formation.

We find that our detailed simulations include most of the physics that is relevant to low-lying Southern Ocean clouds but one particular type of ice formation, called Hallett-Mossop rime splintering, is not active enough. CAM6 and AM4 make too much ice, or glaciate, cumulus clouds. CAM6 has too few cloud droplets and we hypothesize that this is caused by glaciation and by the simulated clouds driving too little turbulent mixing of the atmosphere.

1 Introduction

In the austral summer, highly reflective boundary layer clouds over the Southern Ocean cover nearly two thirds of the 45°S - 65°S latitude band. They increase the albedo of the Earth, reduce sea surface temperatures (SSTs), and moderate global oceanic heat uptake (Roemmich et al., 2015; Hyder et al., 2018). Realistic representation of these clouds in global climate models (GCMs) is vital to simulating the current climate and radiative feedbacks in future, warmer climates. However, GCMs have historically simulated too little low cloud over the Southern Ocean (Trenberth & Fasullo, 2010; Naud et al., 2014).

Insufficient Southern Ocean cloudiness in GCMs has been attributed to the lack of supercooled liquid water in mixed phase clouds within the cold sector of summertime Southern Ocean cyclones (Bodas-Salcedo et al., 2014, 2016). Until recently, almost all GCMs excessively glaciated these clouds (Bodas-Salcedo et al., 2016), which reduces their optical depth (Twomey & Warner, 1967), may reduce their lifetime (Albrecht, 1989), and can lead to overly negative optical depth feedbacks as the simulated Southern Ocean clouds become more liquid-dominated in a warming climate (Tan et al., 2016).

Southern Ocean low clouds form in a unique synoptic environment with distinctive aerosol characteristics. The Southern Hemisphere polar jet generates about 1000 cyclones per year (Yuan et al., 2009), with rapidly evolving extensive low cloud decks in their cold sectors. The absence of land in the Southern Hemisphere extratropics and the strong polar jet isolate the Southern Ocean from continental and anthropogenic sources of dust and aerosol, which affects the nucleation of cloud droplets and ice crystals (Carslaw et al., 2013). As a result, parameterizations of droplet and ice nucleation derived from observations in less pristine locations may not be properly calibrated for Southern Ocean clouds (McCluskey et al., 2018; DeMott et al., 2016).

Southern Ocean boundary layer clouds are commonly mixed-phase, containing supercooled liquid water droplets and smaller concentrations of larger ice particles. Mixed-phase cloud processes, including primary and secondary ice production, the Bergeron-Findeisen mechanism for rapid ice growth (Tan & Storelvmo, 2016), and other mechanisms of cold precipitation formation, are poorly understood compared to warm cloud processes, and climate models have often effectively specified cloud phase as a function of temperature in lieu of realistically representing the melting and freezing of cloud water (McCoy et al., 2016).

For these reasons, much of the GCM cloud physics development going from CMIP5 to CMIP6 targeted reducing the excessive glaciation of Southern Ocean clouds and constraining the low cloud climate feedback. Bodas-Salcedo et al. (2019) found that altering warm rain formation and including turbulent production of liquid water within mixed-phase clouds in HadGEM3-GC3.1 increased cloud liquid water path and reduced the spurious negative feedback associated with mixed-phase low clouds. In the transition from CAM5 to CAM6, Gettelman et al. (2019) found that replacing the ice nucleation and shallow convection schemes with formulations less dependent on temperature accomplished the same thing. While these studies help us understand the controls on liquid water in mixed-phase Southern Ocean clouds within GCMs, a historical dearth of in-situ observations in Southern Ocean clouds has made evaluating these modified microphysics and shallow convection schemes difficult (Tan et al., 2016).

Challenges associated with representing Southern Ocean mixed-phase clouds in GCMs, and evaluating their representation, motivated several recent international efforts to collect measurements of Southern Ocean clouds, aerosols and radiation from ground-based, shipborne and airborne platforms, including several coordinated studies of the region of the Southern Ocean between Australia and Antarctica. Two of these studies were the Clouds, Aerosols, Precipitation Radiation and Atmospheric Composition over the Southern Ocean (CAPRICORN-2) ship campaign and the Southern Ocean Clouds, Radiation and Aerosol Transport Experimental Study (SOCRATES) aircraft campaign, both in January-February 2018. Here, we use a unique multi-sensor suite of SOCRATES observations to build case studies in different types of Southern Ocean cloudy boundary layers and compare them with two GCMs, AM4 and CAM6, run in a nudged-meteorology mode and sampled at the locations and times of the airborne sampling.

Another challenge in representing Southern Ocean boundary layer clouds in GCMs is the complex interplay of large scale synoptic dynamics and smaller scale circulations (such as convection, turbulence, and mesoscale cellularity)(Tomassini et al., 2017), which must be parameterized in GCMs. Large-eddy simulation (LES)- which uses a fine grid to explicitly simulate the cloud-forming eddies over a limited area but must be supplied information about the larger-scale meteorological setting- is a complementary modeling strategy. Thus, we also compare a suitably forced LES with the GCMs and aircraft observations, so that we can fully investigate how model resolution and scale affects Southern Ocean cloud biases. We identify strengths and weaknesses of both the LES and the GCMs as a first step toward improving the representation of Southern Ocean boundary layer clouds in both classes of models.

This work is one of a series of complementary studies using recent observations of Southern Ocean clouds to evaluate nudged GCMs. Zhou et al. 2020 (hereafter Z2020) uses radar reflectivities from aircraft and ship measurements, and in-situ measurements from aircraft to evaluate bulk characteristics of Southern Ocean low and high clouds within CAM6 and AM4. Gettelman et al. 2020 (hereafter G2020) uses SOCRATES in-situ observations to demonstrate that CAM6 maintains liquid water in Southern Ocean mixed-phase clouds more realistically than CAM5, and reproduces the shape of SOCRATES drop and crystal size distributions, but with lower number concentrations than observed. Here, we use cloud resolving simulations to provide process-level explanations for some of the successes and failures of CAM6 and AM4 that are discussed in Z2020 and G2020, especially pertaining to the maintenance of supercooled liquid water and low droplet concentrations in CAM6.

2 Observations

During SOCRATES, the U. S. National Science Foundation Gulfstream-V (G-V) research aircraft, operated by the Research Aviation Facility of the National Center for Atmospheric Research, was based in Hobart, Tasmania, at 43°S, 145°E. The G-V conducted 120 hours of in-situ sampling below, in and above diverse cold-sector Southern Ocean (SO) clouds between 45°S and 62°S during January 15-February 25, 2018. The aircraft instrumentation and flight plans were targeted for constraining cloud-aerosol interactions and mixed-phase microphysics. Cloud probes sized and imaged cloud and precipitation particles and measured condensed cloud mass. Aerosol instruments sized accumulation and coarse mode marine particles and measured concentrations of cloud condensation and ice nuclei. A vertically pointing W-band radar and a High Spectral Resolution Lidar (HSRL) obtained continuous vertical profiles of the cloud and precipitation structures. Unless otherwise noted, all data used here have a time resolution of 1 Hz, corresponding to a horizontal resolution of 120-180 meters, depending on aircraft ground speed.

2.1 Sampling strategy

Research flights typically ferried at an altitude of 6 km to the south end of a target region, launching dropsondes and surveying the underlying clouds with radar and lidar. When approaching the target region, typically at 55-62°S and 135-155°W, the G-V descended to cloud top and reversed direction to conduct sampling 'modules' on the return to Tasmania. Modules ideally consisted of three ten minute level legs – 150 m above the cloud top (above cloud leg), within the cloud layer (in-cloud leg) and 150 meters above the sea surface (below cloud leg) – followed by a sawtooth leg of back-to-back vertical profiles through this entire layer, as shown in Figure 1. In practice, many flights diverged from the ideal sampling strategy in order to sample complex vertical cloud structures, mitigate aircraft icing, or accomplish mission-specific science objectives. Several flights also overflew a measurement site at Macquarie Island (54°S, 157°E) or a research ship, the Australian R/V *Investigator*, which hosted the CAPRICORN2 campaign. The SOCRATES instruments and measurements used in this study are listed in Table 1, where the variable names from the EOL aircraft data files are included in square brackets.

2.2 Vertical velocity variance

An observable measure of turbulence intensity that is predicted by LES and many GCMs is the vertical profile of vertical wind variance, averaged over a sufficiently large horizontal area to fully encompass the most energetic vertical motions. SOCRATES aircraft observations include high-rate 25 Hz vertical wind (w), inferred from multiple pressure measurements and aircraft parameters. The absolute uncertainty in the vertical wind is comparable to typical vertical wind speeds, but variability in the vertical wind is still accurately measured.

Traditionally, vertical velocity variance is estimated from aircraft data using long level legs through relatively homogeneous turbulence. The SOCRATES flights involved extensive profiling, and the boundary-layer cloud often had substantial mesoscale variability and large-scale gradients, so we developed a modified estimation method. We computed the running variance in w ($\sigma^2_{20}[w]$) over a 20 second block centered around the measurement time, which corresponds to a 2.8 km horizontal distance for a typical 140 m s^{-1} G-V ground speed during SOCRATES boundary-layer sampling. This block length is long enough to sample the dominant updraft and downdraft scales and average over aircraft motions, but short enough to resolve fine-scale vertical structures, horizontal trends and mesoscale variability. During SOCRATES, the G-V typically profiled at an ascent/descent rate of 1000 feet per minute. During a 20 second block, its altitude changed by 100 meters, so $\sigma^2_{20}[w]$ encompasses an altitude range much narrower than the typical depth of the boundary layer or a cloud layer.

To correct $\sigma^2_{20}[w]$ for the portion of the true vertical wind variance that occurs on scales larger than 20 seconds, we constructed a power spectrum of w for each below-cloud and in-cloud leg from SOCRATES, and we computed the fraction f_{20} of the vertical wind variance associated with periods greater than 20 seconds. We found that this fraction tends to increase with altitude (z, in meters), so we made the following altitude-dependent correction to $\sigma^2_{20}[w]$ throughout our study to obtain an estimate of the full vertical velocity variance ($\sigma^2[w]$) which can be directly compared with model-derived vertical velocity variance estimates:

$$\sigma^2[w] = \frac{\sigma^2_{20}[w]}{1 - f_{20}}, \tag{1}$$

$$f_{20}(z) = 0.167 + 1.267 \times 10^{-4} \min(z, 2000). \tag{2}$$

2.3 ERA5 reanalysis and its application

We use the fifth generation ECMWF atmospheric reanalysis of the global climate (ERA5) (Hersbach et al., n.d.) for the SOCRATES period to evaluate aircraft measurements and to initialize and force our LES cases. We use hourly pressure level data interpolated onto a horizontal grid of 0.25°×0.25° and 37 pressure levels from its native

137 hybrid sigma/pressure levels and 30 km horizontal grid. Section 4.2 describes how we use ERA5 to set up and force our LES cases.

The G-V radiometric surface temperature brightness (RSTB) can be a valuable proxy for SST when the aircraft is near the sea surface. However, due to calibration drifts, atmospheric absorption, and temperature differences between the instrument and the atmosphere, the RSTB commonly appeared to be offset from the actual SST during SOCRATES. We compare the RSTB with the SST from ERA5, which is strongly constrained with satellite and surface observations. We observe a temperature dependent bias in the RSTB which approaches 2°C at the coldest SSTs (Figure 2a). This discrepancy is larger than the manufacturer's stated temperature dependent uncertainty, which has a maximum of 0.65°C in the SOCRATES dataset. In contrast, the ERA5 SST is unbiased compared with measurements from the *R/V Investigator* from the coinciding CAPRICORN2 experiment (Figure 2b). Thus, we use the ERA5 SST in this work, but we acknowledge that it may not capture mesoscale oceanic eddies which may locally modulate boundary layer stability.

SSTs are tightly coupled to near-surface temperature. We find that ERA5 950-mb temperature compares well with aircraft temperature measurements from vertical profiles, with no mean bias (not shown), lending further credence to the ERA5 SST and near-surface air temperature fields.

3 Stability and cloud morphology regimes

Marine boundary layer clouds are strongly influenced by the air-sea temperature difference. Warmer air traveling over a colder sea surface forms a stable boundary layer with inhibited vertical turbulent mixing and is often accompanied by low-lying cloud layers with different thermodynamic properties than the near-surface air. Colder air traveling over a warmer sea surface drives boundary layer-scale convective eddies, resulting in an unstable and well-mixed boundary layer, usually topped by cumulus and/or stratocumulus clouds. Since stable and unstable boundary layers are both common over the Southern Ocean and involve different physical processes, it is important to test our models in both conditions. We use the SOCRATES observations to investigate how Southern Ocean low cloud morphology varies with boundary layer stability, and then we use this analysis to choose a set of representative cases.

Vertical profiles, typically from sawtooth legs, are selected for this analysis if they extend from below 200 meters altitude up past the bottom of the inversion layer. For profiles with multiple temperature inversions, the aircraft must reach the bottom of the uppermost inversion layer. We estimate the air-sea temperature difference by subtracting the ERA5 SST from the ERA5 2-meter temperature (T_s). If T_s is at least 0.5°C warmer than the SST, we classify the boundary layer as stable. If T_s is at least 0.5°C colder than the SST, we classify the boundary layer as unstable. If the absolute value of the air sea temperature difference is less than 0.5°C, we classify the boundary layer as neutral.

We also classify the cloud morphology sampled within each vertical profile. We smooth the observations by binning the 1 Hz liquid water content (LWC), vertical wind (w), and corrected 20-second running vertical wind variance ($\sigma^2[w]$) into 2 mb pressure bins that span the range from 1050 mb to 400 mb. This binning substantially reduces the noise in the measurements while still resolving sharp temperature inversions. We calculate the bin-medians $\overline{LWC}$, $\overline{w}$ and $\overline{\sigma^2[w]}$. If any pressure bins simultaneously have cloud ($\overline{LWC} >$ 0.01 g kg^{-1}), a strong updraft ($\overline{w} >$ 1 m s^{-1}), and turbulence ($\overline{\sigma^2[w]} >$ 0.1 m^2 s^{-2}), the profile is classified as containing cumulus. If the aircraft profile samples a cumulus-forming environment but does not actually go through a cumulus cloud, then it will not be flagged as containing cumulus.

If cumulus is detected, then we compute a vertically-integrated low cloud fraction for the entire aircraft module containing the vertical profile, using a threshold HSRL backscatter $> 3\times10^{-5} m^{-1} sr^{-1}$ below 4 km elevation as a indicator of the presence of low cloud, as described in Z2020. If the module cloud fraction exceeds 75% then the profile is classified as cumulus rising into stratocumulus; otherwise it is classified as open cell cumulus. If no cumulus is detected, the vertical profile is classified as containing either one stratiform cloud layer or, if there are pressure bins with no liquid water situated between bins containing liquid water, multiple stratiform cloud layers. Figure 3 summarizes this cloud morphology decision tree.

Figure 4 shows the boundary layer stability and cloud morphology for each vertical aircraft leg that profiled the entire boundary layer and sampled cloud. Unstable boundary layers dominate the SOCRATES dataset, as expected for a campaign targeting the cold sectors of cyclones, but stable and neutral boundary layers are each observed in about 20% of SOCRATES profiles. They were most commonly sampled over cold SSTs south of 55 °S. Single layers of stratocumulus cloud are most likely to occur in unstable boundary layers. Multiple stratiform cloud layers are most likely to occur in stable boundary layers, where the top of each cloud layer is typically capped by a temperature inversion. Cumulus rising into stratocumulus occur predominantly in unstable boundary layers, although they can occur in neutral and stable boundary layers, especially in cases where strong meridional winds advect boundary layers into more stable regions but the cloud morphology takes some time to adjust to the reduced forcing at the sea surface. Open cell cumulus were sampled within unstable boundary layers north of 52°S.

3.1 Selection of representative case studies

It is desirable to test our models with a spectrum of cases that span the range of observed boundary layers and cloud morphologies from SOCRATES. The colored rectangles in Figure 4 that are labelled with flight numbers indicate six modules that we have chosen to develop into case studies. These six cases are also described in Table 2. All cases except RF11 are flight modules containing 2 to 4 vertical profiles which observed similar boundary layer and cloud properties throughout the sampling period. Since the selected cases feature similar boundary layer and cloud properties over hundreds of km, it is meaningful to compare them with the ERA5 reanalysis (50 km grid spacing) and the nudged GCMs (100 km grid spacing). Except for RF13, all cases have a 950 mb wind speed U_{950} of 15-20 m/s, which is typical for this part of the Southern Ocean. Cloud top temperatures (T_{top}) of the uppermost sampled cloud layers range from -1.4 °C to -18.2 °C and air sea temperature differences (ΔT) range from -4.05 °C (unstable) to 1.87 °C (stable). All cases feature supercooled liquid water (SLW) dominated clouds with a mixture of frozen and liquid large particles, with the exception of RF13, where most of the cloud is warmer than 0°C. RF11 has limited observations and no complete vertical profiles, but we selected it as the only SOCRATES case featuring open cell cumulus within the Hallett-Mossop temperature range.

4 Models used

4.1 LES Model

Large eddy simulations (LES) model turbulent flows by solving three-dimensional fluid transport equations (including cloud processes, surface fluxes and radiative heating in our case) at a grid scale much smaller than the most energetic eddies but much larger than the scale at which viscosity and molecular diffusivity become important (Smagorinsky, 1963). They include a parameterization of subgrid turbulent eddy effects on the transported fields. LES are useful for studying cloud regimes that strongly interact with turbulent/convective eddies, such as in the Southern Ocean boundary layer. LES of atmospheric boundary layers use domains with a horizontal extent of at least a few times the boundary layer depth. Synoptic dynamics enter into the simulations through the model initialization, advective forcings and other boundary conditions, and target environmental soundings. Internally-generated mesoscale dynamics, often visible in Southern Ocean clouds, can be simulated if the computational domain is sufficiently large (> 50 km) and the simulation is run out at least 12-24 hours.

Our LES study has three major goals: The first is to test whether an LES initialized and forced using either reanalysis or local observations can simulate the typical cloud and boundary layer structures that were observed during SOCRATES. The second is to identify physical processes (e. g. the representation of mixed-phase microphysics) to which the simulated cloud and boundary layer features are sensitive. The third is to compare the LES results, which include a plausible representation of cloud-turbulence interaction, with nudged-hindcast simulations from the CAM6 and AM4 coarse-grid global climate models, run with $\sim$100 km horizontal grid resolution. The GCM boundary layer turbulence and subgrid cloud microphysics parameterizations aim to represent the grid-mean effects of the same processes explicitly simulated by the LES.

We use the System for Atmospheric Modelling (SAM) (Khairoutdinov & Randall, 2003) with Morrison two moment microphysics with graupel (Morrison et al., 2005) (hereafter M2005), the UM5 advection scheme (Yamaguchi et al., 2011) and RRTMG radiation (Mlawer et al., 1997). Of relevance to mixed-phase clouds sampled in SOCRATES, the microphysics scheme includes a parameterization of Hallett-Mossop rime splintering. This scheme allows new ice particles to splinter from graupel and snow at temperatures between -3 and -8°C, when either LWC > 0.5 g kg^{-1} or rain mass > 0.1 g kg^{-1}. Rime splintering is allowed on graupel when its mass exceeds 0.1 g kg^{-1}, and on snow when its mass exceeds 0.1 g kg^{-1}. These thresholds are only rarely surpassed in SOCRATES-sampled low clouds, so unless they are modified, rime splintering is inactive in the cases presented here (Young et al., 2019). We perform sensitivity tests to removing these three thresholds in Section 8.

We specify a uniform cloud droplet concentration for each case, which is equal to the median droplet concentration from the associated vertical aircraft profile. For RF11, which sampled shallow cumulus, we instead use the median droplet concentration from all of the in-situ data from the flight. Table 2 lists the droplet concentrations that are used for each case. Simulations run with interactive aerosol, using a constant bimodal aerosol profile with distribution parameters customized to match each case, simulated very similar number concentrations and produced no detectable changes in cloud macrophysics (not shown).

We choose the domain height to be approximately twice the height of the boundary layer to provide an overlying layer for gravity wave damping. We use a horizontal resolution of 50 m and a square domain with a 12.8 km edge for all cases. We specify a vertical resolution of 10 m in the cloud layer to resolve entrainment, with grid stretching in the overlying atmospheric column. Simulations using 5 m vertical resolution in the cloud layer produce similar results (not shown). We choose the vertical resolution near the surface to be 25 m, within a factor of two of the horizontal resolution, to properly represent near-surface isotropic turbulence and to allow resolved-scale turbulent eddies to efficiently transfer heat and moisture between the near-surface air and the rest of the boundary layer. Two of the cases, RF12 and RF13, with shallow boundary layers, are simulated on a 192-level vertical grid. The other four cases are run on a 320-level vertical grid.

4.2 LES initialization and forcing

SAM uses moist-conserved variables and saturation adjustment to account for the thermodynamics of vapor-liquid phase change. Thus, we use as model input the total specific humidity (q_t, the sum of the mixing ratios of water vapor, q_v, and nonprecipitating cloud condensate, q_c, assumed consistent with observations to be dominated by liquid), and liquid water temperature ($T_L = T - Lq_c/c_p$), computed from either aircraft observations or ERA5 reanalysis.

The Southern Ocean poses unique challenges for our LES framework, due to the strong winds, rapid synoptic variability, and sparsity of detailed observations except by the aircraft itself. Large-scale horizontal advective forcings and vertical motion can cause rapid changes in boundary layer structure at any fixed location, so uncertainty in those inputs can cause a simulation to drift away from reality in as little as an hour. After considerable experimentation, we settled on two LES forcing methodologies with complementary advantages.

An ERA5-based simulation is initialized and forced exclusively with ERA5 reanalysis data and run for 15 hours, reaching the reference time at hour 12. This is a rough analogue to the GCM nudged-hindcast mode. An Obs-based simulation aims to produce a three-dimensional realization of the cloudy marine boundary layer whose domain-mean profiles of temperature, moisture and cloud liquid water match those of the aircraft sounding. Its purpose is to allow comparison of the simulated and observed microphysics without having to account for major differences in the cloud structure. Each simulation is initialized from a single vertical aircraft profile and its horizontal domain-mean q_t and T_L are nudged aggressively ($\tau = 20$ minutes) towards this profile. The solar zenith angle is held constant at the reference time of the case. The simulation is run for 12 hours and hours 10-12 are analyzed here. Most cases reach a steady state profile of cloud liquid water within two hours of simulation but RF09 takes 10 hours to do so. No Obs-based experiment was run for RF11 because the cloud is patchy and we do not have any complete vertical profiles from that flight.

The reference profiles of q_t and T_L are computed from the 2 mb binned aircraft observations. Outside of cloud, q_t is taken as the water vapor specific humidity (q_v) from the VCSEL. In cloud, we estimate the observed cloud condensate q_c from the CDP (assuming the cloud is composed of spherical droplets). We assume that a 2 mb bin is in-cloud if $\overline{LWC} > .005$ g kg^{-1} and we make a linear fit to q_c that extends from the lowermost to the uppermost cloudy bin, for each cloud layer, to make a smoother profile. Although the observed clouds may not be perfectly adiabatic, it is more realistic to nudge q_c to an adiabatic cloud profile with a liquid water path that is close to the observed mean, rather than a profile that has false deviations from an adiabatic profile due to covering a large horizontal area ($\sim$ 10 km). We add the linear fit of q_c to the liquid-saturated water vapor specific humidity (q_s). We use q_s in place of the VCSEL q_v because the LES-simulated q_c is sensitive to any discrepancy from water saturation within clouds in the nudging profile.

Horizontal winds, surface pressure, and SST from ERA5 are used in both the Obs-based and ERA5-based simulations. We estimate the large-scale vertical wind from the

ERA5-reported pressure velocity based on an approximate formula, valid near the surface:

$$\begin{aligned}
\omega(x,y,p,t) &= \frac{D_p p}{Dt} = \frac{D_p p_s}{Dt} + \frac{D_p}{Dt}(p-p_s) \\
&= \frac{D_s p_s}{Dt} + (u-u_s)\frac{\partial p_s}{\partial x} + (v-v_s)\frac{\partial p_s}{\partial y} + \frac{D_p}{Dt}(p-p_s) \\
&\approx \omega_s + w\frac{\partial p}{\partial z}. \qquad (3)
\end{aligned}$$

Here, we use D_p/Dt to be the material derivative at pressure p and D_s/Dt to be the material derivative at the surface pressure p_s. They differ due to the different winds at the two pressures. This is used in the second line above. In the third line, we define $\omega_s = \omega(x, y, p_s, t)$. We also neglect the tendency and horizontal advection of the small quantity $p - p_s$ compared to the vertical advection of p, and we neglect vertical wind shear between p_s and p, which relies on the winds being fairly similar to the surface winds. This approximation ensures that we have no vertical wind at the surface. It breaks down in the upper atmosphere, where the approximation $\omega \approx w\,\partial p/\partial z$ is more robust.

We do an ad-hoc interpolation between these formulas:

$$\omega \approx f(p)\omega_s + w\partial p/\partial z, \qquad (4)$$

where $f(p)$ is a sigmoidal curve that is equal to 1 at the surface and decays to zero at the tropopause (250 mb). Using the hydrostatic approximation to calculate the vertical pressure gradient, we calculate vertical velocity from ERA5 as follows:

$$w \approx -\frac{\omega - \omega_s f(p)}{\rho g}. \qquad (5)$$

We compute geostrophic winds and advective tendencies from ERA5 fields using finite differences. All ERA5 data used as model input has been smoothed over a 1-degree box centered on the model domain. We use the ERA5 reanalysis data on pressure levels for the model input. However, for the Obs-based experiment, we add 2 mb thick levels as needed to resolve all of the observed temperature inversions, and interpolate both observed data and ERA5 reanalysis to the new pressure grid. This ensures that the Obs-based LES includes all of the observed temperature inversions, which are important for the development of stratiform boundary layer clouds. Both LES experiments are nudged towards the ERA5 horizontal winds. The Obs-based simulation uses a wind nudging timescale of 20 minutes and the ERA5-based simulation uses a wind nudging timescale of 1 hour.

4.3 Description of CAM6 and AM4 AGCMs

Our other goal is to evaluate the atmospheric components of two GCMs, CAM6 (Neale et al, The Community Atmosphere Model Version 6, 2020, submitted to *JAMES*) and AM4 (Zhao et al., 2018), that have been run in hindcast mode for the SOCRATES experiment and lightly nudged to reanalysis datasets. Both models use a finite volume dynamical core and comparable grid resolutions.

CAM6 is run on a 0.9°x1.25° latitude/longitude grid with 32 vertical levels. It employs Cloud Layers Unified by Bi-normals (CLUBB) (Guo et al., 2015) to parameterize the turbulence, cloud liquid, and boundary-layer cumulus convection. Its two-moment Morrison-Gettelman microphysics (MG2008) (Morrison & Gettelman, 2008) is analogous to the M2005 scheme in SAM, but optimized for a GCM framework. Unlike M2005, MG2008 doesn't include graupel. However, as shown in Section 5, M2005 does not produce substantial concentrations of graupel in any of the SOCRATES cases. In contrast to the case-specified droplet concentration used for the LES, CAM6 predicts aerosol using the Modal Aerosol Module (MAM4) (Liu et al., 2016), initialized based on climatological profiles in year 2000 from the Coupled Model Intercomparison Project phase 6 (CMIP6) emissions inventory, and explicitly activates cloud droplets. CAM6 is sub sampled along the SOCRATES flight track such that for every ten minutes of observation time, the nearest CAM6 profile to the aircraft location is saved.

AM4 uses a cubed sphere domain with approximately 100 km horizontal resolution and 33 vertical levels. AM4 uses a continuously entrainingdetraining bulk plume based on Bretherton et al. (2004) to represent shallow convection. The microphysics is simpler than in either the LES or CAM6, predicting just four cloud properties including cloud amount, cloud liquid and ice water content, and cloud liquid droplet concentration. AM4 microphysics follows Rotstayn (1997) for hydrometeor mass (which is diagnostic) and cloud fraction and Ming et al. (2007, 2006) for droplet concentration. Cloud droplets are explicitly activated from aerosol, which is predicted based on climatological profiles in year 2016 from the CMIP6 emissions inventory. We use hourly output from AM4 for the Southern Ocean basin.

In CAM6, horizontal winds, temperature, SST and surface pressure are lightly nudged with a timescale $\tau = 24$ hours to the NASA Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2)(Gelaro et al., 2017). AM4 is similarly nudged to ERA5 reanalysis.

5 Model-observation comparisons

We evaluate the ability of the SAM LES and the AM4 and CAM6 GCMs to represent the physical processes that are important for determining the formation, evolution and radiative properties of Southern Ocean boundary layer clouds across our set of cases. For each case, we qualitatively describe the synoptic environment that the clouds have formed and evolved in, using reanalysis and satellite imagery (Section 5.1, Figure 5). We then use observations to evaluate cloud and boundary layer structure, turbulence and cloud microphysics in the models (Sections 5.2-5.5).

5.1 Summary of LES performance for different cloud morphologies

Figure 5 shows a synoptic analysis for each case and compares the cloud morphology simulated by the two different LES experiments with snapshots of the observed clouds from cameras on the aircraft. The top row shows satellites images of visible reflectance along with contours of sea level pressure from ERA5, and the red star within a green circle indicates the location of each case. There are broad correlations between the synoptic environment and the observed cloud morphology. RF01 and RF10 both feature two stratiform cloud layers within westerly flow near 60°S. RF09, RF12 and RF13 are in southwesterly flow, implying that there is more cold advection than in RF01 and RF10, and all three cases feature stratocumulus-topped unstable boundary layers. RF09 has cumulus rising into the stratocumulus layer, due to a greater air-sea temperature gradient (Table 2). RF11 features open cell cumulus in westerly flow nearer to Tasmania, where the warmer sea surface generates strong thermal instability.

Qualitative comparison of cloud morphology between the aircraft snapshots (second row of Figure 5) and the LES experiments (bottom two rows) reveals strengths and weaknesses of the two LES methodologies. For RF01 and RF10, the plane is between the two observed cloud layers during the time of the snapshot. In each case, at least one of the two simulation types captures the cloud and boundary layer structure reasonably well. The Obs-based case is constrained in horizontal mean to have the vertical profile of moisture, temperature and cloud from the observed sounding. Hence it consistently and accurately simulates the cloud morphology of observed stratiform clouds in cases RF01, RF10, RF12 and RF13. However, it cannot simulate rising cumuli in RF09 and is does not capture as much horizontal variability as the ERA5-based simulations. This is because moisture anomalies associated with either rising cumuli, or cloud tops extending above the inversion height specified in the input sounding and into the very dry troposphere ($q_v < 0.5$ g kg^{-1}), are rapidly eroded by the strong nudging.

Since the ERA5-based simulations are not nudged they have more flexibility to simulate an inhomogeneous moisture field, and are therefore better at representing rising cumuli (RF09 and RF11). However, the ERA5-based simulations of the two-layer stratus cases, RF01 and RF10, have trouble simulating more than one cloud layer, because these thin cloud layers are tied to fine scale features in the input temperature and humidity soundings that are not resolved by ERA5 reanalysis.

Figure 6 shows a comparison of simulated reflectivities from the Obs-based LES experiments and observed reflectivities from the G-V cloud radar, for all six cases. The yellow line indicates the vertical aircraft profile that the Obs-based LES is nudged to and the plots show the entire aircraft modules that are used to evaluate the LES and GCMs throughout this section. The cloud morphology is usually consistent throughout the module, with considerable mesoscale variability in the observed reflectivities. Since the small-domain LES experiments cannot capture this mesoscale variability, their reflectivities tend to fall inside a narrow window within the range sampled by the flight modules. For example, in RF01, a multi-layer stratus case, the simulated clouds are lightly precipitating everywhere, consistent with the observed reflectivities at 02:10 UTC. The ERA5-based and Obs-based simulations generally simulate similar ranges of reflectivities, suggesting that the cloud microphysics is not strongly tied to the cloud morphology. The LES does not simulate the highest observed reflectivities in cases RF01, RF11, RF12 and RF13. This may be due to a deficiency of large particles, a lack of mesoscale variability, or a combination thereof. We will show in Sections 5.3 and 5.5 that there are too few large particles in the LES in cases RF11 and RF12.

5.2 Observational case-by-case model evaluation methodology

We start by evaluating the temperature and moisture profiles in the models, including the location of temperature inversions, since those properties determine boundary layer mixing and the amount of moisture available to the clouds. Since the studied clouds are dominated by SLW, we compare profiles of cloud fraction and in-cloud LWC between the models and observations to evaluate the cloud macrophysics. In cases where the models simulate a substantial amount of ice mass, we show the ice water content (IWC) as well, but there are no direct measurements of IWC that can be used to evaluate the models.

Turbulent eddies, including convection, are vital for cloud formation, boundary layer structure, and vertical mixing. The vertical structure of turbulence within SOCRATES boundary layers is determined by surface heat fluxes, near surface wind shear, and clouds. SOCRATES boundary layers are usually characterized by decoupled turbulence profiles with distinct peaks near the sea surface and within each cloud layer. Multi-layer stratus clouds and cumulus are associated with stronger decoupling. We compare turbulence between the observations and models using profiles of the vertical velocity variance. For the LES, we estimate the vertical wind variance by adding the resolved vertical wind variance and 2/3 of the subgrid scale turbulent kinetic energy (i. e. equipartitioning of subgrid turbulent kinetic energy between coordinate directions). In general, the resolved contribution dominates the subgrid contribution. For CAM6, we further examine how the turbulent structure of the boundary layer affects simulated cloud condensation nuclei (CCN) and droplet concentrations. As a reminder, the LES uses fixed droplet concentrations, which are specified in Table 2.

Cloud microphysics influences cloud lifetime, cloud radiative effects and precipitation. To evaluate simulated cloud microphysics, we compare particle size distributions (PSDs) between observations and models. We have separate PSDs for each hydrometeor from the model output, but it is much more challenging to classify the observations by hydrometeor phase. A synthesis of data from four G-V particle probes in Z2020 suggests that in most supercooled boundary-layer clouds observed in SOCRATES at temperatures of -5 to -25°C, the largest particles (diameter $D > 200$ μm), when present, were predominantly frozen (graupel and snow), the smallest particles ($D < 50$ μm) were predominantly liquid, and midsize particles ($50 < D < 200$ μm) could be either drizzle or ice.

5.3 Single-layer stratocumulus cases RF12 and RF13

RF12 and RF13 sampled extensive single-layer stratocumulus decks within southwesterly cyclonic flow, as shown in the satellite images in Figures 5e-f. Figure 7 compares the observed temperature, q_v, and LWC with the LES experiments (top row), the GCMs, and ERA5 reanalysis (bottom row). Cloud fraction is compared between the low and high resolution models, but there is no comparable observed variable. The dashed black lines show the profile that the obs-based case is nudged towards. All solid lines indicate medians and all shaded areas indicate the 10th to 90th percentile. The observations are binned into 2 mb pressure bins spanning the range from 1050 mb to 400 mb, before the statistics are computed. Although the Obs-based LES is nudged to a single profile, we evaluate it using the entire aircraft module, to show that the chosen aircraft profile and the Obs-based LES are representative of a larger area. The black dashed line is an interpolation between an aircraft profile and the ERA5 reanalysis so it is sometimes outside of the 10th to 90th percentile range of the observations.

The thermodynamic profiles for both cases in Figures 7 show well-mixed boundary layers topped by approximately adiabatic stratocumulus cloud layers. For RF12, the cloud layer occupies the Hallett-Mossop temperature range, (-3 to -8 °C, indicated with yellow shading) whereas it is almost entirely above freezing for RF13, making an interesting microphysical comparison. By construction, the Obs-based LES closely reproduces the observed boundary layer properties and cloud macrophysics for both cases. This is also true for the RF12 ERA5-based LES (Figure 7a), suggesting ERA5 is representing the synoptic environment of that case well. For RF13, the ERA5-based run correctly simulates a well-mixed stratocumulus-topped boundary layer that is deeper than the Obs-based case but still within the range of observations. However, it is too dry near the surface, and the cloud is too thin. The ERA5-based LES cloud morphology resembles that of the ERA5 reanalysis (green line in Figure 7d), suggesting the biases may originate from the input soundings, rather than from the LES physics. The ERA5-based LES develops severe cold biases in the free troposphere in both cases, which, through entrainment, leads to modest cold biases throughout the boundary layer. We interpret these as artifacts of the LES response to strong horizontal warm advection at the inversion level, as discussed in the appendix.

The two GCMs, CAM6 and AM4, and ERA5 reanalysis, reproduce the observed supercooled liquid water dominated stratocumulus layers for both RF12 and RF13. AM4 has a lower cloud fraction than observed. All three models, and the GCMs in particular, simulate clouds with too low peak LWC. In the GCMs, the clouds also extend slightly too high because their capping inversion is smeared out. Both of these biases are expected consequences of the coarse vertical resolution of the GCMs. The tendency of the GCM clouds to be too deep was also noted in Z2020 for flight RF12.

The left column of Figure 8 compares $\sigma^2[w]$ between the observations, the LES experiments and CAM6. AM4 and ERA5 do not output turbulence variables. Observed $\sigma^2[w]$ profiles show enhanced turbulence within the cloud layer above uniform weaker turbulence in the subcloud layer. For RF12, the ERA5-based LES features stronger turbulence than the Obs-based LES for RF12, possibly related to the free tropospheric temperature biases, but both experiments are within the range of the observations. CAM6 produces too much turbulence near the surface but too little turbulence in the stratocumulus layer. For RF13, the ERA5-based LES underpredicts cloud-driven turbulence, likely due to the simulated cloud being too thin, and has too much turbulence near the surface. CAM6 produces a well-mixed turbulence profile, missing the in-cloud enhancement. Overall, the LES captures the vertical features of the turbulence profile better than CAM6.

The middle column of Figure 8 compares droplet concentration between the LES observations, CAM6 and AM4, and the right column compares the CCN concentration at a supersaturation of 0.5% from CAM6 with the observed concentration from the UHSAS of particles greater than 0.1 μm. AM4 does not output CCN concentration. We use the large particles from the UHSAS as a proxy for CCN because, although there were two CCN counters on the G/V, one was scanning through a wide range of supersaturations at all times, and the other had frequent problems leading to missing data.

The CAM6 droplet concentration is too low in both the RF12 and RF13 stratocumulus, and in most other SOCRATES cases. The reasons for this bias seem to be regime-dependent, but are particularly inobvious in this regime. AM4 droplet concentration is comparable to CAM6 in these two cases, but tends to be higher and more consistent with observations in other cases that we'll show shortly. CAM6 simulates realistic CCN concentrations in both cases, suggesting that in this regime, too few CCN do not lead to too few droplets.

Figure 9 compares average in-cloud PSDs between the observations (using the G-V CDP and 2DS instruments to span the full size range), the Obs-based LES and CAM6. This complements results in G2020, which show PSDs for the entirety of SOCRATES but not separated by cloud regimes. The Obs-based LES is used for the comparison because, of the two LES experiments, its bulk cloud properties have better agreement with the observations. The AM4 climate model and ERA5 reanalysis use one-moment microphysics schemes so they are left out of this comparison.

We develop a robust cloud flag using 10 Hz LWC data from the CDP. 1 Hz aircraft data is considered in-cloud when all 10 subsamples of CDP data have LWC > 0.01 g m^{-3}. In-cloud PSDs are averaged over the entire flight module associated with each case. The CAM6 PSD is averaged over all in-cloud grid cells overlapping the flight module. For both CAM6 and the LES, grid cells are used if they have in-cloud LWC > 0.01 g m^{-3}, consistent with the processing of the observations.

In both single-layer stratocumulus cases, the LES and CAM6 simulate qualitatively similar PSDs, despite having large discrepancies in cloud macrophysics. The LES and CAM6 have too little drizzle (diameters around ~ 100 μm). This bias may be partly an artifact of representing the droplet and rain populations as lognormal distributions, and may be improved by adding a a third class of liquid particles to the bulk microphysics scheme to represent drizzle (Sant et al., 2015, 2013).

Additionally, both the LES and CAM6 fail to simulate the large particle mode ($\sim$ 300 μm - 1 mm) in RF12, which we assume is primarily snow and graupel. As discussed in G2020 and Z2020, CAM6 does not have this bias in other cases or in comparisons with all SOCRATES data. We hypothesize that this may be due to insufficient Hallett-Mossop rime splintering in the LES and CAM6 simulations. Indeed, there is no rime splintering in the LES because the simulated hydrometeor masses do not exceed the thresholds specified in the M2005 Hallet-Mossop parameterization described in Section 4.1. In Section 8, we remove these thresholds and find that the Obs-based LES is able to reproduce the observed large particle mode. The MG2008 microphysics in CAM6 includes a parameterization of Hallett-Mossop rime splintering without mass thresholds; it is unclear if it is active in this case because process rates were not included in the model output. G2020 found that turning Hallett-Mossop rime splintering off in the nudged CAM6 hindcast had little effect on simulated liquid and ice water paths and droplet concentrations for the SOCRATES time period, and we found that the cloud macrophysics and microphysics for this case were unchanged in that simulation.

5.4 Two-layer stratus cases RF01 and RF10

RF01 and RF10 both sampled westerly flow behind cold fronts in regions of large scale ascent near 60°S (Figures 5a,c). RF01 includes four vertical profiles through two stratus layers featuring a remarkably consistent cloud morphology throughout the sampling period (Figure 6a), with strong mesoscale variability in the top cloud layer. The aircraft sampled more variable cloud morphology in RF10, with discontinuities in both the upper and lower cloud layers (Figure 6c).

Figure 10 compares the observed thermodynamic profiles with the LES experiments, ERA5 reanalysis, CAM6 and AM4. In the case of RF10, ERA5 has substantial ice water content (IWC) and a dotted line has been added to the profile of LWC to indicate IWC. These are particularly challenging cases to model. Both cases have strong temperature inversions capping the upper cloud layer and decoupled moisture profiles with strong gradients in humidity above each cloud layer. RF01 has a small temperature inversion atop the lower cloud layer, but RF10 just has a layer of slightly reduced lapse rate there. The LES captures the observed cloud morphology only if it simulates the observed temperature inversions. For RF01, the ERA5-based LES simulates the uppermost observed temperature inversion and a robust upper cloud layer (Figure 10a), and has very sparse condensation at the location of the observed lower cloud layer. We examined the temperature and humidity profiles from all $0.25° \times 0.25°$ ERA5 reanalysis grid cells that have been averaged together to make the smooth 1-degree input for the ERA5-based simulation, on both pressure and model levels. None of the horizontal grid cells captured the observed temperature inversions or decoupled moisture profile.

For RF10, the ERA5-based LES has a strong inversion at 900 mb, and only simulates the lower cloud layer (Figure 10b). The Obs-based LES, on the other hand, simulates two temperature inversions and two robust cloud layers in both cases. The Obs-based LES, on the other hand, simulates two temperature inversions and two robust cloud layers in both cases.

For this discussion, because our focus is low cloud formation, we define the top of the boundary layer as the location where the humidity decreases to typical free tropospheric values (< 0.5 g kg^{-1}), which is the uppermost inversion in these cases. The strongly decoupled moisture profiles suggest that there is little or no mixing between the two cloud layers, such that the atmosphere above the lower cloud is not turbulently interacting with the surface, and hence would not be included in a classical definition of the boundary layer.

ERA5, CAM6 and AM4, which have coarser vertical grid resolution than the LES, cannot consistently represent the observed temperature inversions, boundary layer humidity profiles and cloud morphologies (Figures 10c,d). For RF01, ERA5 broadly matches the observed thermodynamic structure and simulates scattered thin cloud throughout the lower troposphere. Both GCMs have a moist bias throughout the boundary layer, and CAM6 has a warm bias as well, which results in simulated cloud layers that are too deep; CAM6 also has unrealistically high in-cloud LWCs in the lower parts of the cloud (Figure 10c). For RF10, AM4 simulates just the lower cloud layer, which is consistent with the range of observations, and ERA5 and CAM6 have two distinct cloud layers. Both GCMs have a cold bias throughout the boundary layer. ERA5 has partially glaciated the top of the cloud but still maintains supercooled liquid water throughout the boundary layer.

The left column of Figure 11 compares turbulence between the observations, the LES experiments and CAM6. The observed turbulence in both cases is enhanced within the upper cloud layer and exhibits a smaller peak near the surface, below 950 mb. Both LES experiments capture both peaks (even if the cloud layers are not in the correct places). CAM6 captures the near surface shear-driven turbulence in both cases but entirely misses the cloud-driven turbulence, like for the RF12 stratocumulus case, but with larger biases here.

The middle and right columns of Figure 11 evaluate simulated droplet concentrations in both GCMs, and simulated CCN in CAM6, respectively. For RF01, both GCMs achieve realistic droplet concentrations within the lower cloud layer for RF01 but exhibit unrealistic drop-offs above it. For RF10, AM4 produces somewhat too many droplets in the lower cloud layer (the only one that it simulates). CAM6 produces a realistic droplet concentration in the lower cloud layer but has too few droplets in the upper cloud layer.

CAM6 underestimates CCN throughout the boundary layer by nearly 50% in RF01 but has has a realistic CCN profile in RF10. In both cases, CAM6 turbulence, CCN concentration and droplet concentration all peak at the surface. Turbulence can influence droplet concentration both by modifying aerosol transport and the efficiency of aerosol activation within the cloud layer. Since the low bias in the droplet concentration profile is much more pronounced than in the CCN profile, it is plausible that CAM6 is activating too few CCN due to insufficient turbulence at the base of the upper cloud layer.

Figure 12 compares module-average in-cloud PSDs between the observations, the Obs-based LES and CAM6. As in RF12 and RF13, the LES under-predicts drizzle in both cases. CAM6 under-predicts drizzle in RF10. It is difficult to compare snow concentrations between simulations and observations due to poor counting statistics for the low concentrations of large particles that were observed.

5.5 Cumuliform clouds

Figure 13 compares thermodynamic profiles between observations, the LES experiments, ERA5 reanalysis, CAM6 and AM4. These cases contain a substantial amount of ice and dotted lines showing IWC have been added to the plot of LWC. Unlike for earlier cases, LWC and IWC are plotted on a log scale. For the LES, autoconversion of ice to snow is very efficient, so the profile of IWC for the LES is computed from the snow mass only.

For RF11, (Figures 13b,d), all 1 Hz LWC observations are plotted as black dots because the clouds were sampled only at a few heights.

For RF09, the ERA5-based LES has too deep a cumulus cloud layer (Figure 13a) with too much LWC. This bias may partly be inherited from ERA5 reanalysis, which is also too dry above 900 mb and simulates too high a cumulus top at 700 mb (Figure 13c). The Obs-based LES simulates a more realistic profile of LWC in the upper part of the cloud but creates a spurious stratiform cloud layer at 840 mb, where the observed sounding sampled a cumulus cloud, due to the strong nudging. Both LES experiments simulate SLW-dominated clouds with substantial ice mass ($\sim 10\%$ of the condensed mass).

For RF09, all low resolution models simulate deep clouds, like the ERA5-based LES, with varying degrees of glaciation (Figure 13c). CAM6 and AM4 have negligible fractions of LWC throughout the cloud, whereas ERA5 maintains a substantial amount of SLW between 850 and 900 mb.

For RF11, the ERA5-based LES simulates a range of LWC which agrees very well with the observed range of LWCs at the heights where we have observations, and a negligible IWC (Figure 13b). On the other hand, CAM6 and ERA5 both have glaciated cloud tops and CAM6 has too little condensed cloud mass (Figure 13d). CAM6 also has too high of a cloud fraction, implying that it simulates a thin, homogeneous cloud deck, instead of a patchy field of thicker clouds. AM4 and ERA5 simulate more realistic cloud fractions with in-cloud LWCs consistent with the range of observations.

In both cases, the deep convection scheme in CAM6, which does not use the MG2008 microphysics, turns on. This scheme also turns on in RF10 but does not lead to excessive glaciation. Glaciation in CAM6 may be substantially reduced if MG2008 microphysics is run within the deep convection scheme.

The left panels of Figure 14 compares the observed and simulated turbulence profiles for these two cases. The Obs-based LES represents the cloud-driven turbulence well in the stratocumulus layer for RF09, but underestimates it below that layer because it does not simulate the rising cumuli that are responsible for generating the turbulence (Figure 14a). The ERA5-based LES captures the turbulence peaks driven by rising cumuli and the stratocumulus deck. The ERA5-based LES also reproduces the multi-peaked observed turbulence for RF11 (Figure 14b). CAM6 simulates a realistic turbulence profile for RF11 and for the lower boundary layer in RF09, but misses the stratocumulus-driven turbulence in RF09.

As shown in the center and right panels of Figure 14, CAM6 has realistic or excessive CCN concentrations but too few droplets in both cases. AM4 simulates droplet concentrations within the observed range in both cases. In this case, the low droplet concentrations in CAM6 may be driven by excessive glaciation.

Figure 15 compares in-cloud PSDs between the observations, the ERA5-based LES and CAM6. We use the ERA5-based LES to evaluate the microphysics in these cases because there is no Obs-based LES experiment for RF11, and because the Obs-based LES does not simulate the rising cumuli in RF09. An in-cloud LWC threshold is used for all simulations, to be consistent with the CDP, so entirely glaciated grid cells from CAM6 are not included in the PSDs.

The LES and CAM6 both reproduce the observed mid-size ($\sim$ 100 μm) and large ($\sim$ 300 μm - 1 mm) particle modes skillfully for RF09, but it is likely that glaciated grid cells in CAM6 have higher concentrations of frozen particles. Z2020 noted deficient droplet concentrations in CAM6 for RF09 and also found that CAM6 precipitated too frequently in that case.

For RF11, CAM6 has a negligible droplet concentration (too low to appear on the plot). The LES and CAM6 are both deficient in large particles, although CAM6 simulates higher concentrations than than LES. RF11 has the highest observed concentration of large particles of the cases presented in this study and likely has substantial IWC but there is no reliable way to quantify IWC from aircraft measurements. The negligible IWC in the LES (Figure 13b) is likely unrealistic and a result of deficient production of large frozen particles. The cumuli in this case partially overlap the Hallett-Mossop temperature range. The lack of large particles, which was also seen in RF12, may stem from the parameterization of secondary ice production within M2005 being inactive, as we will now show.

6 Sensitivity to primary and secondary ice production

In this section, we perform two microphysical sensitivity tests on case RF12 using the Obs-based LES. This single-layer stratocumulus case is attractive because it is within the Hallett-Mossop temperature range (for which M2005 has an ice multiplication parameterization), and the cloud geometry is well simulated. For convenience, the upper and lower panels of Fig. 16a repeat the PSD and synthetic radar reflectivity for this baseline simulation. In contrast to the observations, the PSD shows almost no large particles, and the reflectivity is correspondingly weak.

First, we test the sensitivity of the RF12 simulation to turning off the ice microphysics in M2005 (no primary ice production). This has very little impact on the PSDs of the droplets and rain but slightly decreases the already low synthetic reflectivities (Figure 16b)a.

Second, we remove all of the mass thresholds (described in Section 4.1) from the Hallett-Mossop scheme to increase the production of ice and snow. Since the graupel concentrations are so low in this case, all of the splintering occurs on snow particles. Turning on the Hallett-Mossop processes enables the LES to skillfully reproduce the observed large particle mode. It also initiates precipitation in the LES and drastically increases the simulated reflectivities(Figure 16c), which now resemble the highest reflectivities from the observations (Figure 6c). Ice and snow concentrations increase 100-fold, while there is a modest reduction of cloud and drizzle droplets (Figure 16c) and a 30% decrease in the liquid water path (Figure 17), part of which is due to the Hallett-Mossop process reducing the number of cloudy grid cells and is not captured within the in-cloud PSDs. Figure 17 also shows that it takes six hours to spin this process up due to the initial lack of snow to initiate the ice multiplication.

7 Summary of model successes and biases

The SOCRATES campaign sampled stratiform and cumiliform clouds within thermally unstable, neutral and stable boundary layers. We develop six case studies from SOCRATES data to test the ability of the SAM LES, CAM6 and AM4 to represent diverse SLW-dominated Southern Ocean boundary layer clouds.

LES: The Obs-based LES forcing methodology works well for simulating multiple stratiform cloud layers (RF01 and RF10) and for simulating the vertical structure of the turbulence observed in the stratiform cases (RF01, RF10, RF12 and RF13). This is likely because the thin cloud layers and boundary layer decoupling are maintained by small scale features in the observed temperature and humidity profiles that are not captured in the ERA5 reanalysis data. The ERA5-based LES works well for simulating cumiliform clouds (RF09 and RF11), for which strong nudging to a single thermodynamic profile cannot capture the horizontal inhomogeneity of the cloud field.

The LES has too few drizzle drops (100 - 300 μm diameter) in stratiform cases (RF01, RF10, RF12 and RF13) but this is likely due to the constraints imposed by the bulk microphysical scheme, rather than a bias in the model physics. The LES produces too little snow and graupel with larger diameters of 300 μm - 1 mm in cases RF11 and RF12. This bias can be rectified in case RF12 by removing thresholds in the M2005 microphysical parameterization that inactivate Hallett-Mossop rime-splintering.

GCMs: The two nudged GCMs, CAM6 and AM4, correctly simulate SLW-dominated clouds in all stratiform cases (RF01, RF10, RF12 and RF13) and achieve the most realistic cloud morphology for the two stratocumulus cases (RF12 and RF13), although

AM4 has a lower cloud fraction than observed. Both GCMs fail to simulate the observed temperature inversions capping the cloud layers in the two stable cases (RF01 and RF10).

CAM6 simulates excessively glaciated clouds for the two cumuliform cases, RF09 and RF11, and over predicts cloud fraction in both cases. Excessive glaciation is likely caused by the deep convection scheme activating. AM4 glaciates the rising cumuli and overlying stratocumulus clouds in RF09 but captures the open cell cumuli well for RF11.

CAM6 simulates a droplet concentration that is too low in all six cases, but the bias is larger in the two cumiliform cases (RF09 and RF11). AM4 has realistic droplet concentrations in most cases but has a high bias for RF10 and a low bias for RF13.

CAM6 provided turbulence statistics, CCN concentrations, and cloud PSDs that we compared with SOCRATES observations. CAM6 simulated the turbulence in cumulus layers well. However, in the three stable and weakly unstable boundary layer cases (RF01, RF10, RF12), CAM6 has excessive near surface turbulence and very little cloud-driven turbulence. CAM6 also misses the stratocumulus-driven turbulence observed in RF09. CAM6 has CCN concentrations at least as large as observed in all cases except RF01, so lack of CCN cannot explain the systematically low droplet concentrations. We hypothesize that too little cloud-driven turbulence and excessive glaciation contribute to this bias in CAM6.

8 Conclusions

Improving projections of future climate necessitates constraining cloud-aerosol interactions and climate feedbacks associated with extratropical clouds (Zelinka et al., 2020). As GCM development targets this goal, recent observations of Southern Ocean clouds make it possible to continuously evaluate simulated clouds and cloud processes against the real world. The SOCRATES dataset provides simultaneous measurements of aerosols, and microphysical and macrophysical cloud properties, useful for evaluating the whole spectrum of physics schemes associated with cloud formation in GCMs and process models, such as LES, that can help guide GCM development. Section 7 summarizes our findings about mixed-phase cloud microphysics and turbulence structure from a comprehensive comparison of two GCMs and an LES with six SOCRATES-observed cases sampling three boundary layer cloud regimes: single-layer stratocumulus, two-layer stratus, and cumulus with or without overlying stratocumulus. We encourage other modelling teams to use the SOCRATES dataset and the LES cases presented here, to methodically evaluate and improve simulations of Southern Ocean clouds from a process-level perspective.

Acknowledgments

The authors acknowledge funding from U. S. National Science Foundation (NSF) grants AGS-1660604 (UW and NCAR) and an NSF Graduate Student Research Fellowship to R. Atlas. The data used in this study rely on a collaborative effort by a broad, dedicated SOCRATES team led by Greg McFarquhar. We thank all the SOCRATES scientists and the NCAR Earth Observing Laboratory (EOL) for collecting the data and helping us interpret it. We additionally thank Xiaoli Zhou for providing a lidar-based cloud fraction and corrected radar data from the aircraft, Johannes Mohrmann for his help with combining aircraft observations with ERA5 reanalysis data to make SAM LES input, Jorgen Jensen for his help in using high frequency measurements of liquid water content and vertical wind to develop a cloud flag and turbulence proxy, Wei Wu for processing the 2DS measurements, Joe Finlon and Greg McFarquhar for helping to interpret the 2DS measurements, Julie Haggerty for her help in understanding the uncertainties in the RSTB measurement, and Marat Khairoutdinov for developing and maintaining the SAM LES. All aircraft data is publicly available at `https://data.eol.ucar.edu/master_lists/generated/socrates/`. SAM LES input for the cases described here is publicly available at `https://atmos.uw.edu/~ratlas/SOCRATES_LES_cases.html`.

9 Appendix

The ERA5-based LES featured drastic cold biases in the free troposphere, just above the boundary layer, for the two stratocumulus cases, RF12 and RF13 (Figures 7a,b). Southern Ocean stratocumulus cases are often associated with strong, rapidly evolving inversions. Figure 18a shows an example of this (note solar noon is at 2.5 UTC). An inversion develops at 800 mb at the beginning of the simulation and by 2 UTC, the inversion height has decreased to 850 mb. Although this is a gradual and continuous process in ERA5, in which the inversion becomes stretched over two layers and eventually drops down to the lower layer, it is translated into a discrete, discontinuous process when it is interpolated onto the high resolution LES grid. In the LES, the inversion abruptly drops several vertical layers at about -4 UTC, due to strong horizontal advection, and a new inversion develops at 850 mb. However, the LES has no way to erode the pre-existing inversion at 800 mb and ends up with two inversions, which develop a large cold bias between them (figure 18b). This mid-tropospheric cold bias is improved when the model is nudged to ERA5 reanalysis data and is able to erode the upper inversion. However, because the inversion in ERA5 is sloped and exists partly within the cloud layer, even a modest nudging timescale (τ=12 hours) substantially reduces the cloud thickness by drying out the top of the cloud layer. Using low resolution reanalysis data as input to an LES in a synoptically active region like the Southern Ocean can lead to errors in the representation of stratocumulus-topped boundary layers associated with sharp temperature inversions. Nested simulations using strong nudging only at the edges of the domain may be more suitable for simulating stratocumulus-topped boundary layers over the Southern Ocean.

References

Albrecht, B. A. (1989). Aerosols, cloud microphysics, and fractional cloudiness. *Science*, *245*(4923), 1227-1230. doi: 10.1126/science.245.4923.1227

Bodas-Salcedo, A., Hill, P. G., Furtado, K., Williams, K. D., Field, P. R., Manners, J. C., … Kato, S. (2016). Large contribution of supercooled liquid clouds to the solar radiation budget of the southern ocean. *Journal of Climate*, *29*(11), 4213-4228. doi: 10.1175/jcli-d-15-0564.1

Bodas-Salcedo, A., Mulcahy, J. P., Andrews, T., Williams, K. D., Ringer, M. A., Field, P. R., & Elsaesser, G. S. (2019). Strong dependence of atmospheric feedbacks on mixed-phase microphysics and aerosol-cloud interactions in hadgem3. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1735-1758. doi: 10.1029/2019ms001688

Bodas-Salcedo, A., Williams, K. D., Ringer, M. A., Beau, I., Cole, J. N. S., Dufresne, J. L., … Yokohata, T. (2014). Origins of the solar radiation biases over the southern ocean in cfmip2 models. *Journal of Climate*, *27*(1), 41-56. doi: 10.1175/jcli-d-13-00169.1

Bretherton, C., McCaa, J., & 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. doi: 10.1175/1520-0493(2004)132(0864:ANPFSC)2.0.CO;2

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., ... Pierce, J. R. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature*, *503*(7474), 67-+. doi: 10.1038/nature12674

DeMott, P. J., Hill, T. C. J., McCluskey, C. S., Prather, K. A., Collins, D. B., Sullivan, R. C., ... Franc, G. D. (2016). Sea spray aerosol as a unique source of ice nucleating particles. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5797-5803. doi: 10.1073/pnas.1514034112

Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., ... Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, *30*(14), 5419-5454. doi: 10.1175/jcli-d-16-0758.1

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., ... Mills, M. J. (2019). High climate sensitivity in the community earth system model version 2 (cesm2). *Geophysical Research Letters*, *46*(14), 8329-8337. doi: 10.1029/2019gl083978

Guo, H., Golaz, J. ., Donner, L. J., Wyman, B., Zhao, M., & Ginoux, P. (2015). Clubb as a unified cloud parameterization: Opportunities and challenges. *Geophysical Research Letters*, *42*(11), 4540-4547. doi: 10.1002/2015GL063672

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Hornyi, A., Muoz-Sabater, J., ... Thpaut, J.-N. (n.d.). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *n/a*(n/a). doi: 10.1002/qj.3803

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., ... Belcher, S. E. (2018). Critical southern ocean climate model biases traced to atmospheric model cloud errors. *Nature Communications*, *9*. doi: 10.1038/s41467-018-05634-2

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

Liu, X., Ma, P. L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., ... Rasch, P. J. (2016). Description and evaluation of a new four-mode version of the modal aerosol module (mam4) within version 5.3 of the community atmosphere model. *Geoscientific Model Development*, *9*(2), 505. doi:

10.5194/gmd-9-505-2016

McCluskey, C. S., Hill, T. C. J., Humphries, R. S., Rauker, A. M., Moreau, S., Strutton, P. G., ... DeMott, P. J. (2018). Observations of ice nucleating particles over southern ocean waters. *Geophysical Research Letters*, *45*(21), 11989-11997. doi: 10.1029/2018gl079981

McCoy, D. T., Tan, I., Hartmann, D. L., Zelinka, M. D., & Storelvmo, T. (2016). On the relationships among cloud cover, mixed-phase partitioning, and planetary albedo in gcms. *Journal of Advances in Modeling Earth Systems*, *8*(2), 650-668. doi: 10.1002/2015ms000589

Ming, Y., Ramaswamy, V., Donner, L. J., & Phillips, V. (2006). A new parameterization of cloud droplet activation applicable to general circulation models. *Journal Of Atmospheric Sciences*, *63*(4), 1348-1356. doi: 10.1175/JAS3686.1

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.(author abstract). *Journal of the Atmospheric Sciences*, *64*(4), 1189. doi: 10.1175/JAS3874.1

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: Rrtm, a validated correlated-k model for the longwave. *Journal of Geophysical Research-Atmospheres*, *102*(D14), 16663-16682. doi: 10.1029/97jd00237

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., & 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), 3642-3654,3656-3659. doi: 10.1175/2008JCLI2105.1

Naud, C. M., Booth, J. F., & Del Genio, A. D. (2014). Evaluation of era-interim and merra cloudiness in the southern ocean. *Journal of Climate*, *27*(5), 2109-2124. doi: 10.1175/jcli-d-13-00432.1

Roemmich, D., Church, J., Gilson, J., Monselesan, D., Sutton, P., & Wijffels, S.

(2015). Unabated planetary warming and its ocean structure since 2006. *Nature Climate Change*, *5*(3), 240-245. doi: 10.1038/nclimate2513

Rotstayn, L. D. (1997). A physically based scheme for the treatment of stratiform clouds and precipitation in largescale models. i: Description and evaluation of the microphysical processes. *Quarterly Journal of the Royal Meteorological Society*, *123*(541), 1227-1282. doi: 10.1002/qj.49712354106

Sant, V., Lohmann, U., & Seifert, A. (2013). Performance of a triclass parameterization for the collision-coalescence process in shallow clouds [Journal Article]. *Journal of the Atmospheric Sciences*, *70*(6), 1744-1767. Retrieved from `<GotoISI>://WOS:000319668400014` doi: 10.1175/jas-d-12-0154.1

Sant, V., Posselt, R., & Lohmann, U. (2015). Prognostic precipitation with three liquid water classes in the echam5-ham gcm [Journal Article]. *Atmospheric Chemistry and Physics*, *15*(15), 8717-8738. Retrieved from `<GotoISI>://WOS:000359949700010` doi: 10.5194/acp-15-8717-2015

Smagorinsky, J. (1963). General circulation experiments with the primitive equations: I. the basic experiment. *Monthly weather review*, *91*(3), 99-164.

Tan, I., & Storelvmo, T. (2016). Sensitivity study on the influence of cloud microphysical parameters on mixed-phase cloud thermodynamic phase partitioning in cam5. *Journal of the Atmospheric Sciences*, *73*(2), 709-728. doi: 10.1175/jas-d-15-0152.1

Tan, I., Storelvmo, T., & Zelinka, M. D. (2016). Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science*, *352*(6282), 224-227. doi: 10.1126/science.aad5300

Tomassini, L., Field, P. R., Honnert, R., Malardel, S., McTaggart-Cowan, R., Saitou, K., … Seifert, A. (2017). The "grey zone" cold air outbreak global model intercomparison: A cross evaluation using large-eddy simulations. *Journal of Advances in Modeling Earth Systems*, *9*(1), 39-64. doi: 10.1002/2016ms000822

Trenberth, K. E., & Fasullo, J. T. (2010). Simulation of present-day and twenty-first-century energy budgets of the southern oceans. *Journal of Climate*, *23*(2), 440-454. doi: 10.1175/2009jcli3152.1

Twomey, S., & Warner, J. (1967). Comparison of measurements of cloud droplets and cloud nuclei. *Journal of the Atmospheric Sciences*, *24*(6), 702-703. doi: 10.1175/1520-0469(1967)024⟨0702:comocd⟩2.0.co;2

Yamaguchi, T., Randall, D. A., & Khairoutdinov, M. F. (2011). Cloud modeling tests of the ultimate-macho scalar advection scheme. *Monthly Weather Review*, *139*(10), 3248-3264. doi: 10.1175/mwr-d-10-05044.1

Young, G., Lachlan-Cope, T., O'Shea, S. J., Dearden, C., Listowski, C., Bower, K. N., ... Gallagher, M. W. (2019). Radiative effects of secondary ice enhancement in coastal antarctic clouds. *Geophysical Research Letters*, *46*(4), 2312-2321. doi: 10.1029/2018gl080551

Yuan, X. J., Patoux, J., & Li, C. H. (2009). Satellite-based midlatitude cyclone statistics over the southern ocean: 2. tracks and surface fluxes. *Journal of Geophysical Research-Atmospheres*, *114*. doi: 10.1029/2008jd010874

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., ... Taylor, K. E. (2020). Causes of higher climate sensitivity in cmip6 models [Journal Article]. *Geophysical Research Letters*, *47*(1). Retrieved from `<GotoISI>://WOS:000513983400038` doi: 10.1029/2019gl085782

Zhao, M., Golaz, J. ., Held, I. M., Guo, H., Balaji, V., Benson, R., ... Xiang, B. (2018). The gfdl global atmosphere and land model am4.0/lm4.0: 2. model description, sensitivity studies, and tuning strategies. *Journal of Advances in Modeling Earth Systems*, *10*(3), 735-769. doi: 10.1002/2017MS001209

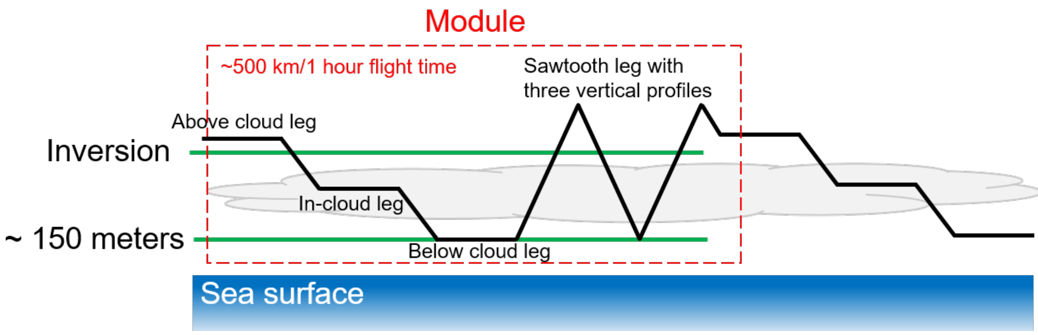


Figure 1. Ideal SOCRATES flight module (dashed red box) with a below cloud leg, in-cloud leg, above cloud leg, and sawtooth leg comprised of profiles that extend from the subcloud layer to the free troposphere.

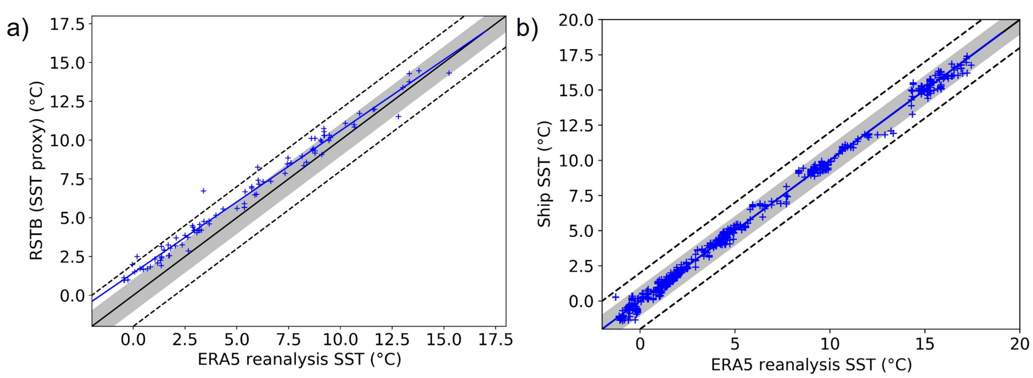


Figure 2. ERA5 SST scattered with a) radiometric surface temperature brightness (RSTB) from the aircraft and b) SST from the *R/V Investigator*. Blue crosses are the data, the blue line is the best fit line and the black line is the 1-to-1 line. Within the shaded region, the difference between the SST estimates is less than 1° and within the dashed lines, it is less than 2°.

Table 1. SOCRATES aircraft measurements used in the study

	Instrument	**Measurement**
Atmospheric Parameters	HARCO heated total air temperature sensors	Temperature [ATX]
	Parascientific Sensor, Model 1000	Pressure [PSXC]
	Vertical-Cavity Surface-Emitting Laser (VCSEL)	Water Vapor [MR]
Bulk cloud properties	Cloud Droplet Probe (CDP)	Liquid Water Content [PLWCD_RWIO]
	HIAPER Cloud Radar	Reflectivity
	High Spectral Resolution Lidar (HSRL)	Scattering and depolarization
	Wintronics KT19.85 Radiometer	Radiometric surface temperature brightness (SST proxy) [RSTB]
Microphysical Properties	Ultra-high sensitivity aerosol spectrometer (UHSAS)	Large aerosol concentration [CONCU100_LWII]
	Cloud Droplet Probe (CDP)	Size distribution (2-50 μm) [CCDP_RWIO]
	Two-Dimensional Stereo Probe (2-DS)	Size distribution (25 μm - 1 mm)
Dynamics	Pressure ports, inertial reference systems and GPS	Horizontal and vertical wind [UIC,VIC,WIC]

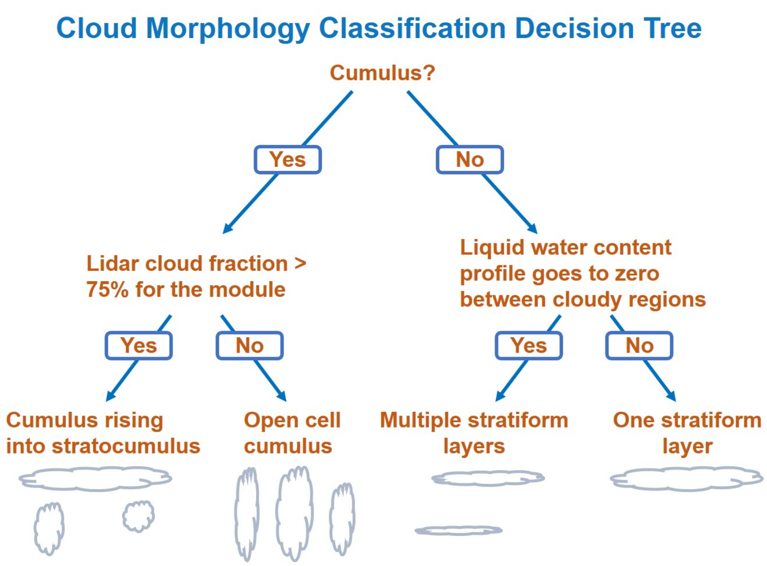


Figure 3. Decision tree showing how aircraft measurements of liquid water content, vertical wind, vertical wind variance, and lidar data from the HSRL, are used to classify the cloud morphology sampled in every vertical profile from SOCRATES, into one of four categories.

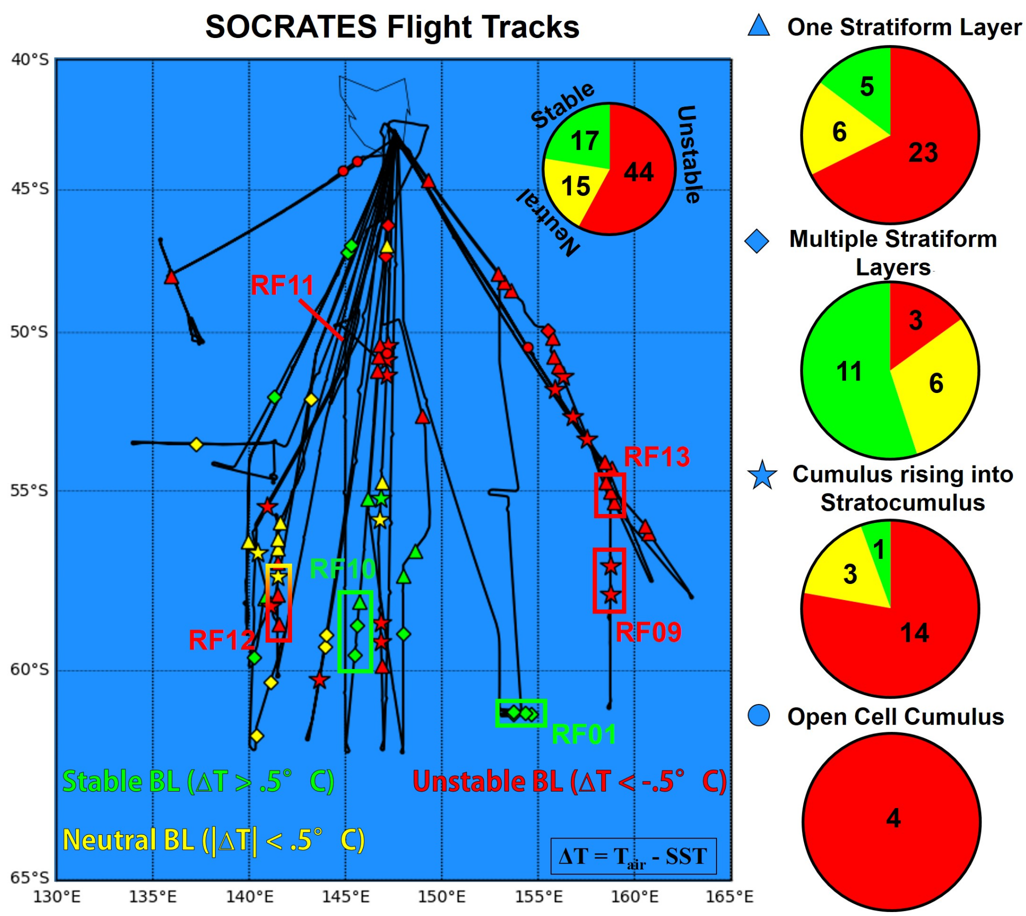


Figure 4. Map of the Southern Ocean between Antarctica and Tasmania with SOCRATES flight tracks (black lines) and symbols showing the location of the vertical aircraft profiles from SOCRATES that profiled the entire boundary layer and sampled cloud. The colors of the symbols represent boundary layer stability and their shapes represent cloud morphology. Rectangles highlight the modules that have been developed into case studies. The pie charts on the right show the frequency of each combination of boundary layer stability and cloud morphology.

Table 2. SOCRATES Cases

Flight	**Date/Time (2018/UTC)**	**Reference Time**	**Location**	**Stability**	**Cloud Regime**	$\mathbf{T_{top}}$ **(°C)**	$\mathbf{N_d}$ $\mathbf{(cm^{-3})}$	$\mathbf{U_{950}}$ $\mathbf{(m\ s^{-1})}$	$\mathbf{\Delta T}$ **(°C)**
RF01	Jan 16 1:50-2:45	2:30	61°S, 154.25 °E	Stable	Two St Layers	-3.3/-11.8	75	19	1.44
RF09	Feb 5 2:40-4:00	3:30	58°S, 158.75 °E	Unstable	Cu under Sc	-18.2	190	15	-4.05
RF10	Feb 8 0:45-1:25	1:00	59.5°S, 145.5 °E	Stable	Two St Layers	.6/-13.4	55	17	1.87
RF11	Feb 17 3:00-5:00	4:00	51°S, 144.5 °E	Unstable	Open-Cell Cu	N/A	115	17	-1.48
RF12	Feb 18 3:25-4:10	4:00	57°S, 141.5 °E	Neutral	Sc	-7.5	210	17	-0.70
RF13	Feb 20 2:50-3:30	3:00	54.75°S, 158.5 °E	Unstable	Sc	-1.4	180	8	-0.95

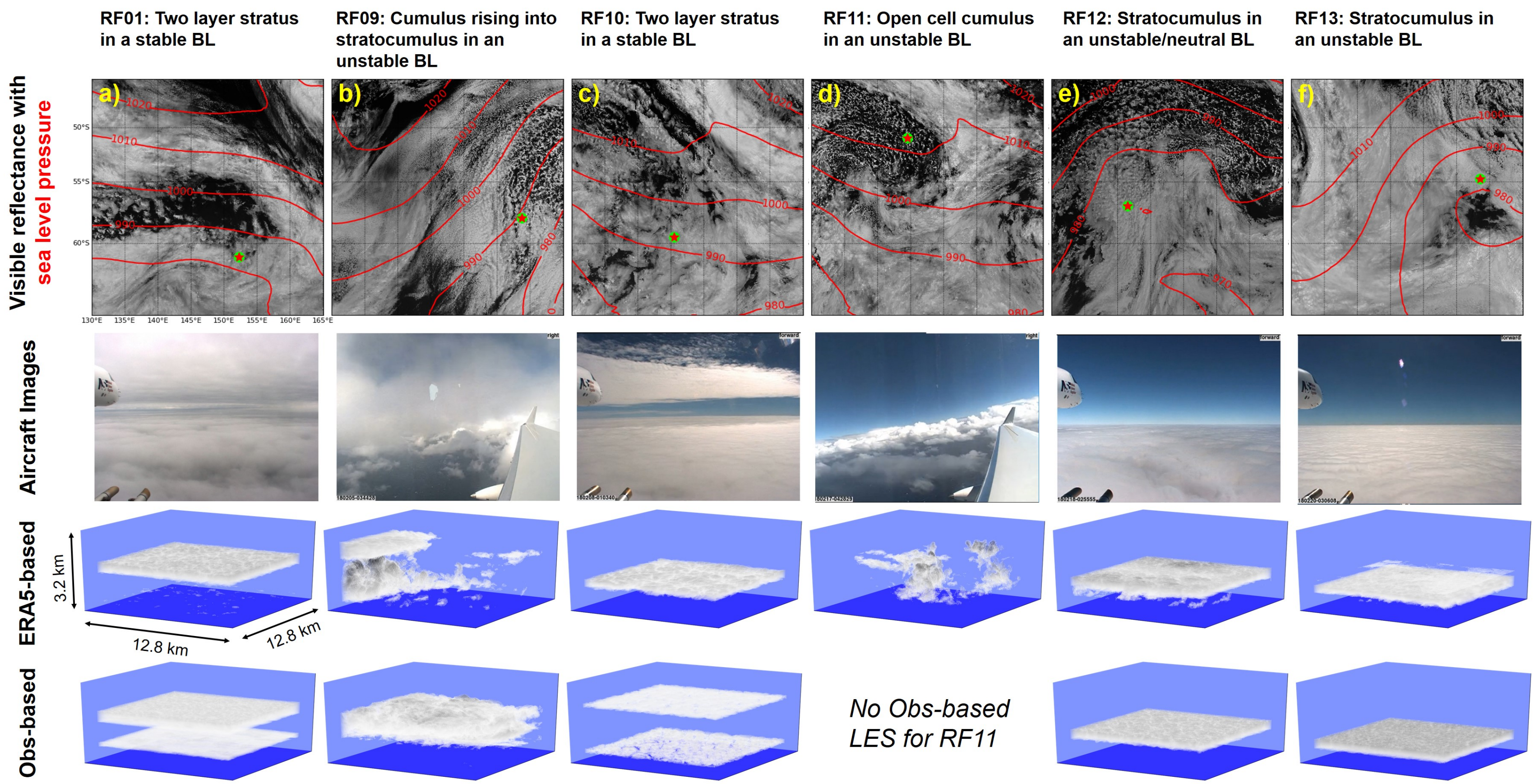


Figure 5. a) Visible reflectance from Himawari with contours of sea level pressure from ERA5 reanalysis. The stars inscribed in the circles show the locations of the cases b) Snapshots during the flight module corresponding to each case from the aircraft's front facing, right side or left side camera and c) LWC at hour 12 of the ERA5-based simulations, which correspond to the reference times of each case d) LWC at hour 12 of the Obs-based simulations, at which point all of the simulations have been in a steady state for 2 hours

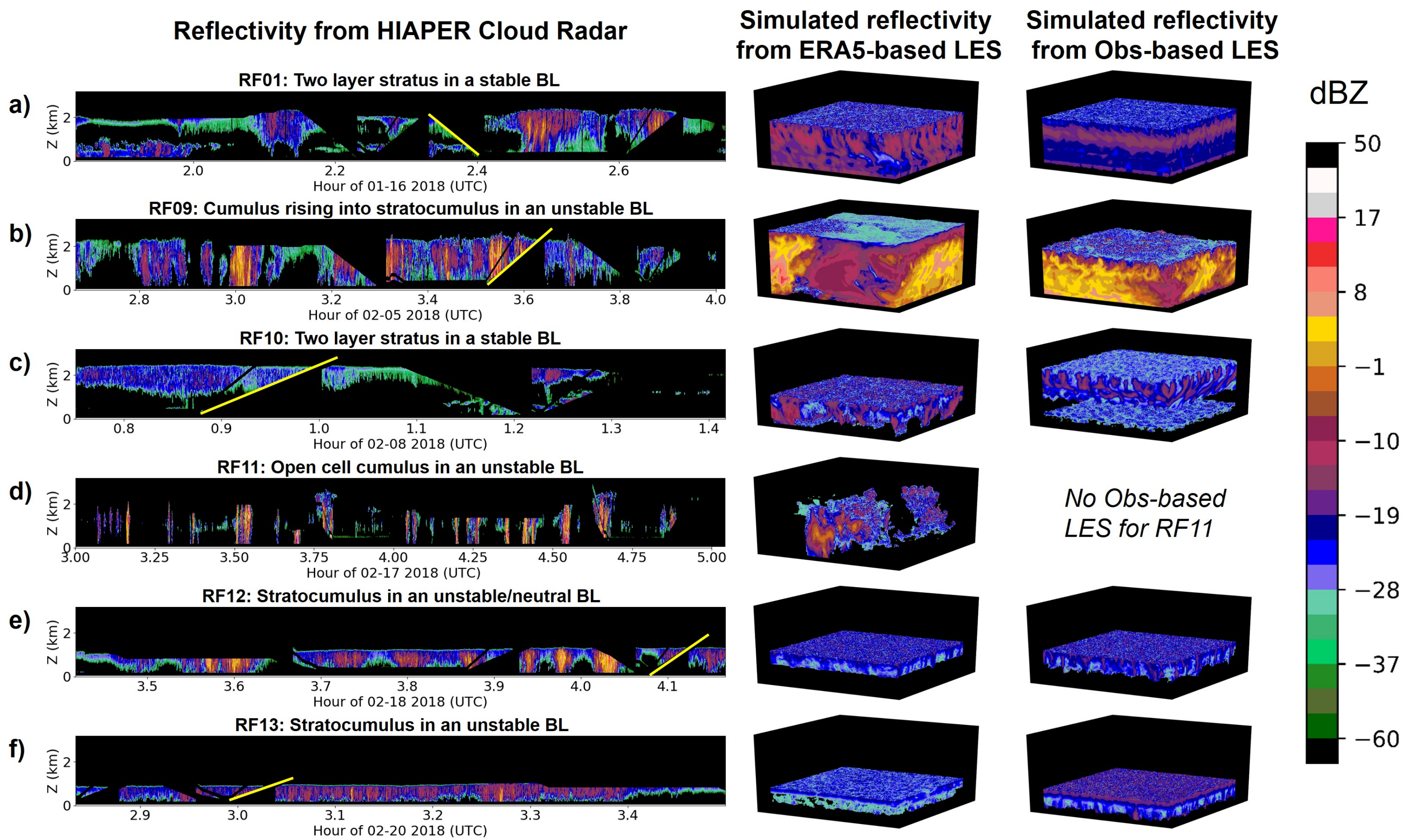


Figure 6. Observed reflectivities (left column) and simulated reflectivities from the ERA5 and obs-based LES runs (right columns) for a) RF01, b) RF09, c) RF10, d) RF11, e) RF12 and f) RF13. Yellow lines indicate the vertical profiles to which the Obs-based LES simulations are nudged. Small LES domains do not capture mesoscale variability and are best compared with the observed reflectivity profiles along the yellow lines.

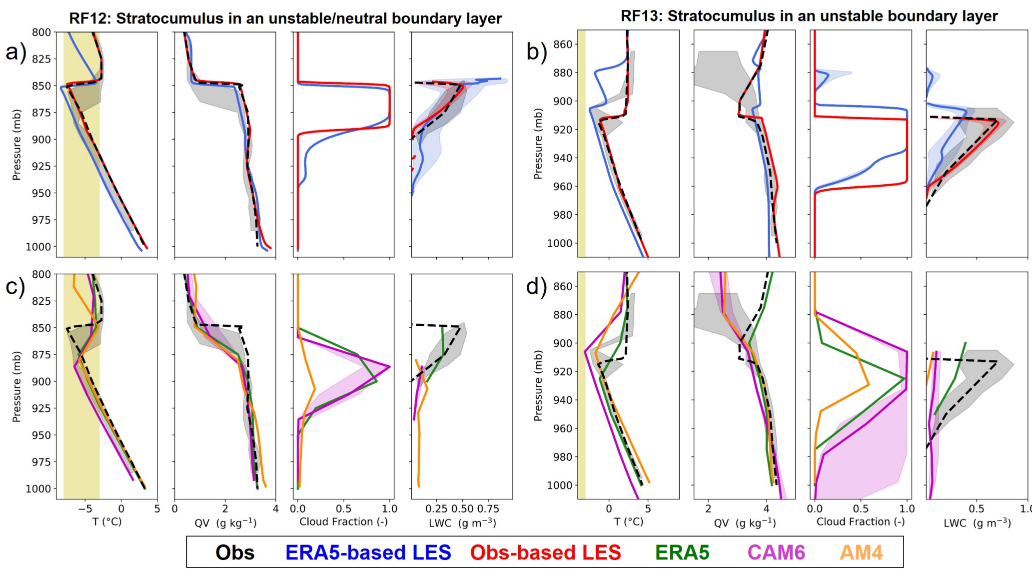


Figure 7. Top row: Temperature, q_t and LWC from observations and SAM LES for the two stratocumulus cases, a) RF12 and b) RF13. Bottom row: The same variables from observations, compared with ERA5, CAM6 and AM4, for the two stratocumulus cases, c) RF12 and d) RF13. The dashed black line shows the profile that the Obs-based LES is nudged towards. The shaded yellow region on the temperature plot indicates the range in which Hallett-Mossop rime splintering occurs. Grey, blue, red and purple shading in the top row shows the 10th to 90th percentile.

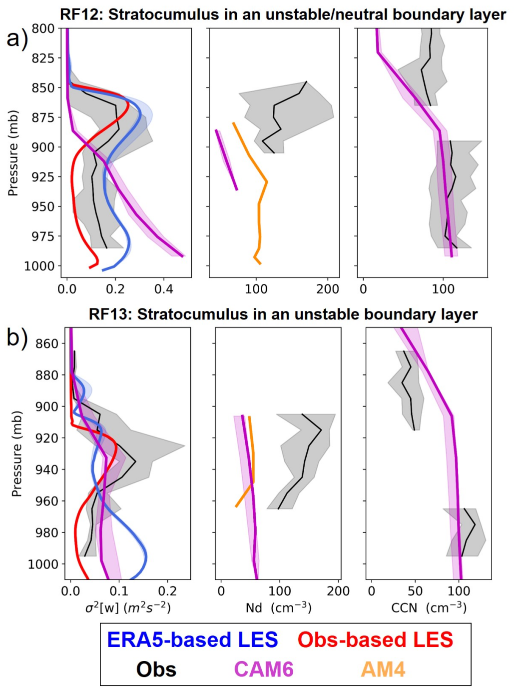


Figure 8. Left column: Profiles of vertical velocity variance $\sigma^2[w]$ from G-V profiles, SAM LES, and CAM6, for the two stratocumulus cases, a) RF12 and b) RF13. Center column: Number concentration from the CDP, and the two GCMs, CAM6 and AM4. Right column: CCN concentration from CAM6 at a supersaturation of .5% and the observed number of aerosols greater than .1 μm, as measured by the UHSAS. All solid lines indicate medians and all shaded regions indicate the 10th to 90th percentile.

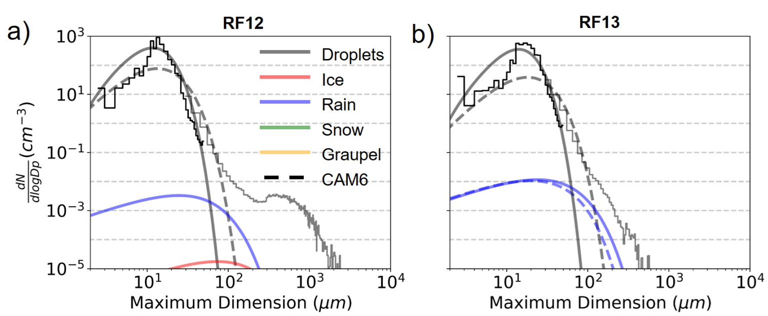


Figure 9. Particle size distributions from the CDP (thin black line) and 2DS (thin grey line), the Obs-based LES (thick solid lines) and CAM6 (thick dashed lines), for the two stratocumulus cases, a) RF12 and b) RF13. For these cases, graupel (included in the LES but not CAM6) has concentrations too low to show up in the plot, but it will be evident in other cases. LES PSDs are from the Obs-based experiments.

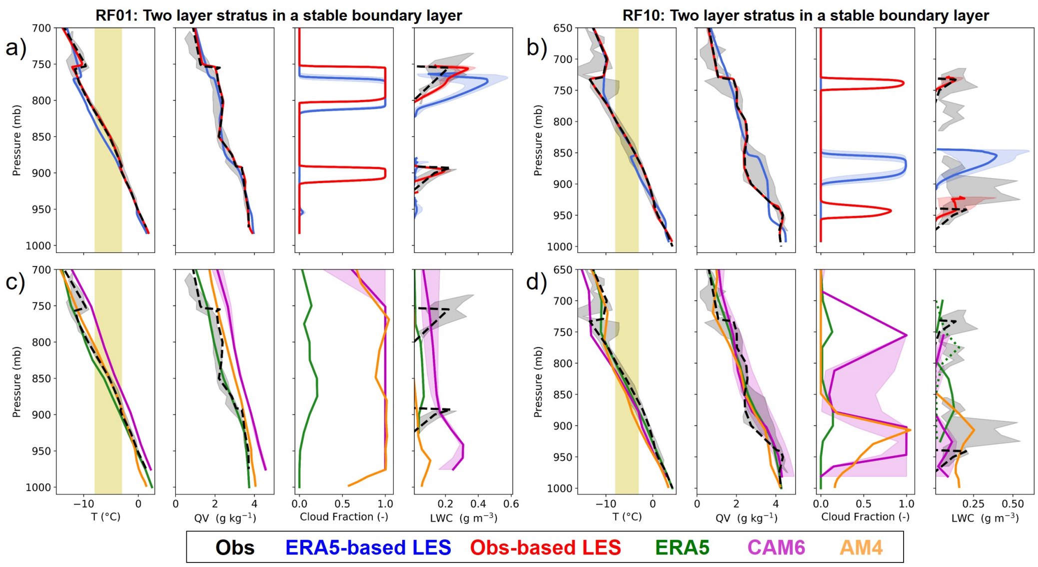


Figure 10. Same as Figure 7 but for the two-layer stratus cases, a) RF01 and b) RF10. Dotted lines indicating IWC have been added to the profile of LWC for the low resolution models, for RF10 (rightmost plots).

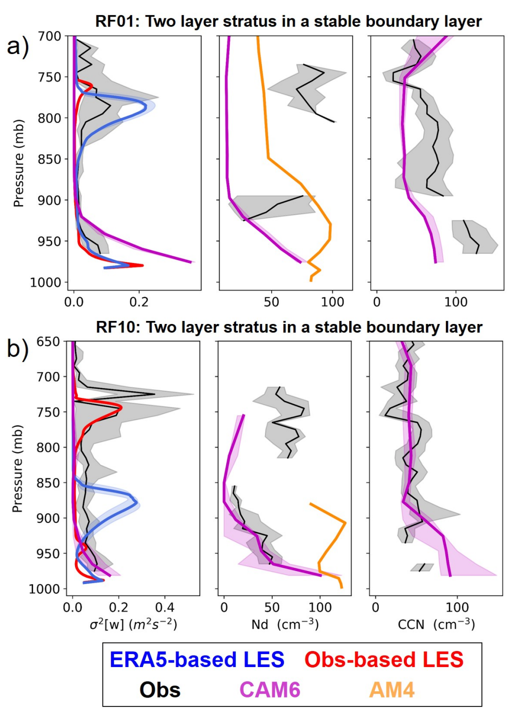


Figure 11. Same as Figure 8 but for the two-layer stratus cases, a) RF01 and b) RF10.

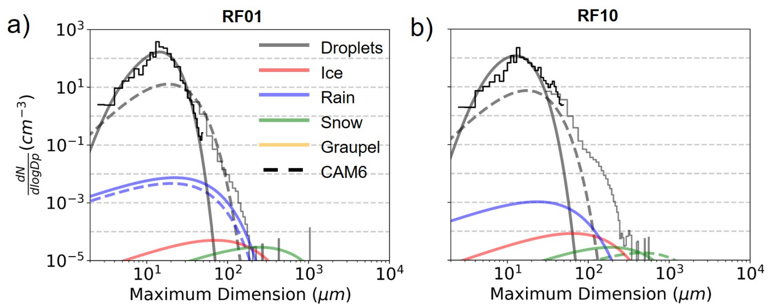


Figure 12. Same as figure 9 but for the two-layer stratus cases, a) RF01 and b) RF10. LES PSDs are from the Obs-based experiments.

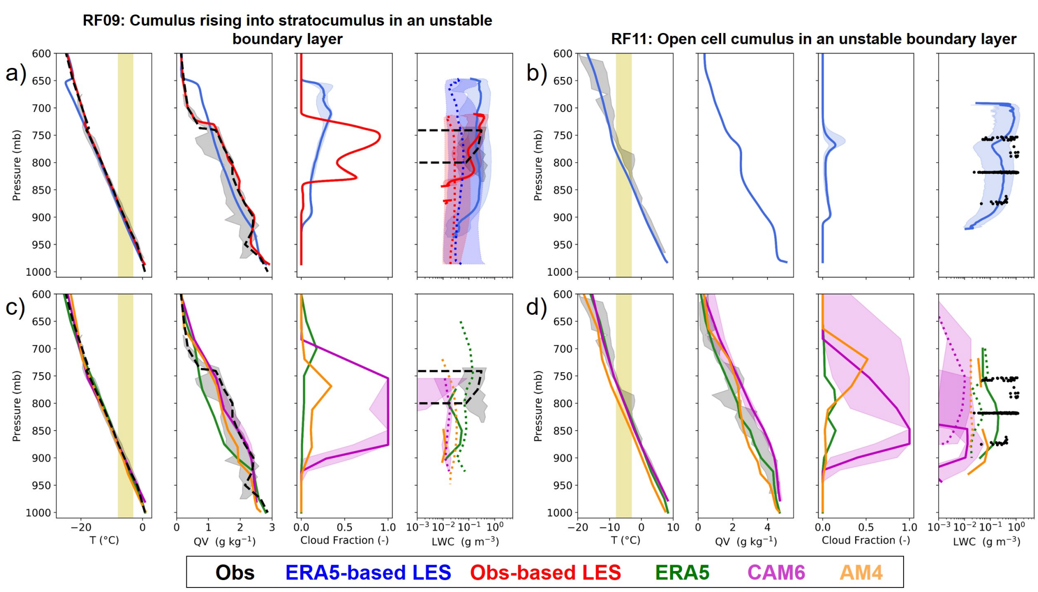


Figure 13. Same as figure 7 but for the two cumuliform cases in unstable boundary layers, a) RF09 b) RF11. LWC (solid line) and IWC (dotted line) are both shown for all simulations and are plotted on a log scale. There are no direct measurements of IWC. For RF11, all 1 Hz in-cloud LWC is plotted because in-cloud measurements are sparse and there are no vertical profiles through cloud from this flight.

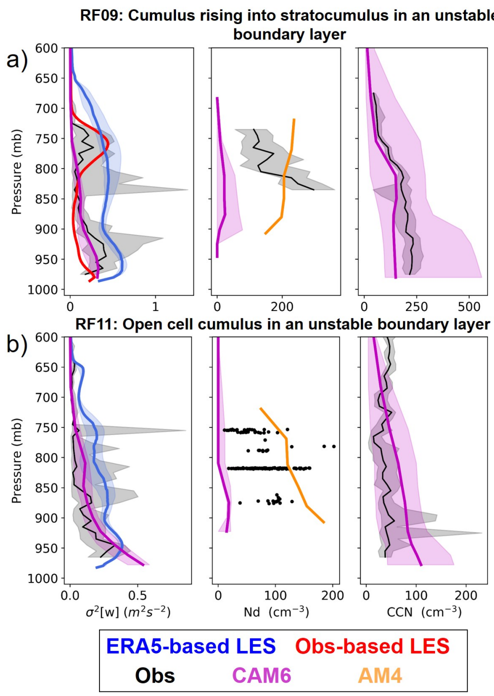


Figure 14. Same as Figure 8 but for the two cumuliform cases, a) RF09 and b) RF11.

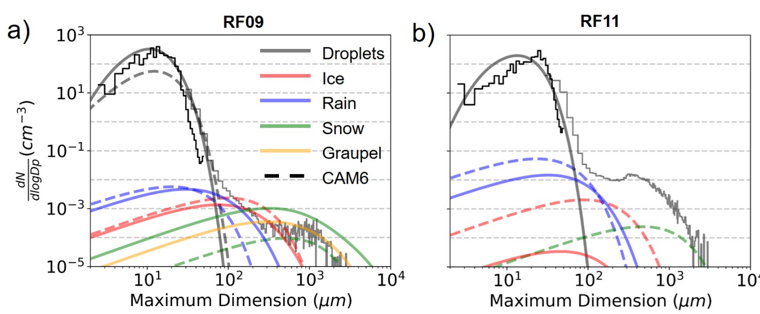


Figure 15. Same as Figure 9 but for the two cumuliform cases, a) RF09 and b) RF11. LES PSDs are from the ERA5-based experiments.

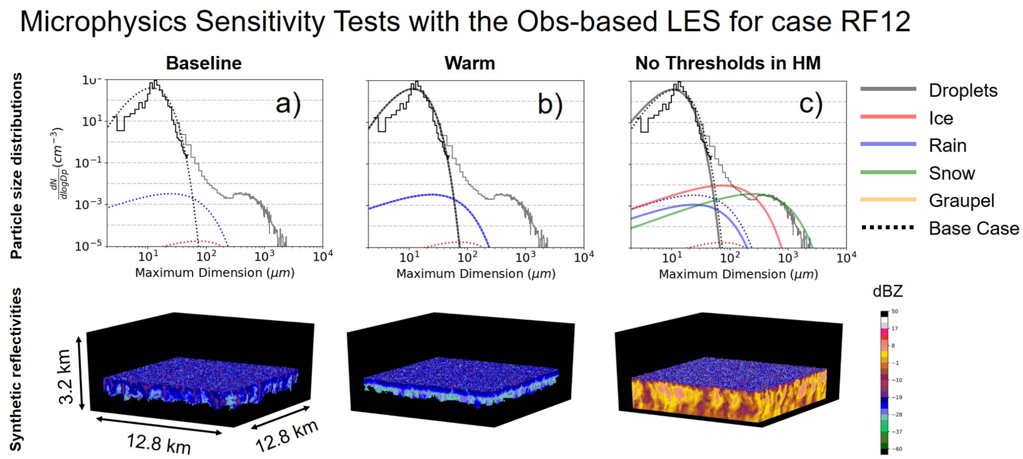


Figure 16. PSDs (top row) and synthetic reflectivities (bottom row) are shown for the baseline Obs-based simulation of RF12 (a) and two microphysics sensitivity tests (b and c). The dashed lines in all plots show the baseline PSDs.

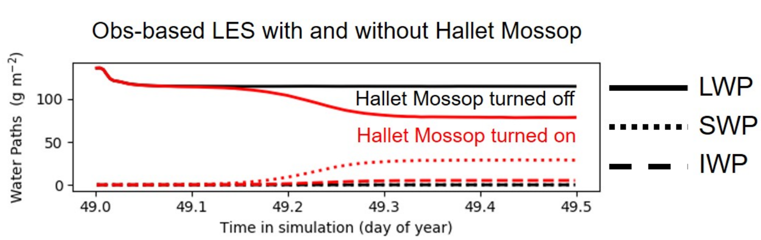


Figure 17. A time series of water paths for liquid, snow and ice are shown for the baseline Obs-based simulation (black) and the simulation with the thresholds in the Hallett-Mossop parameterization removed (red).

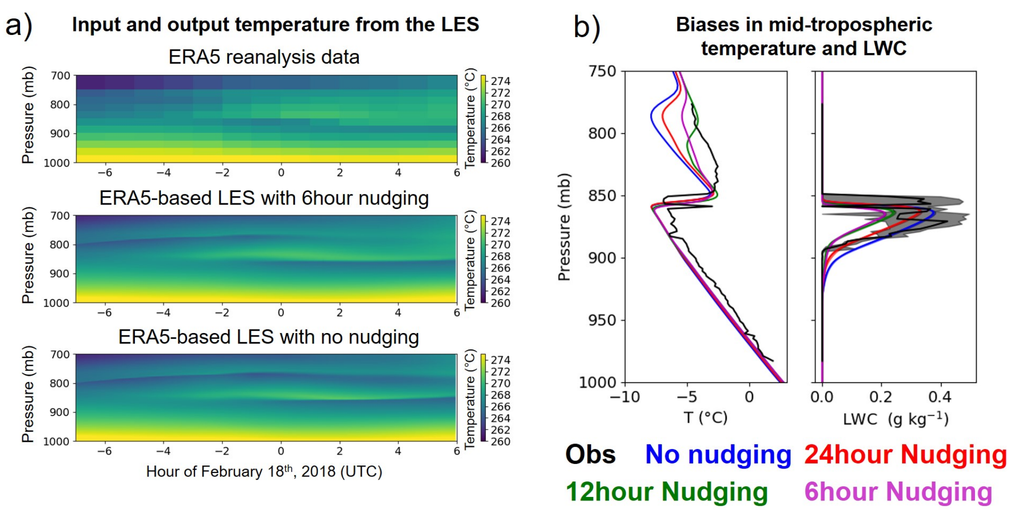


Figure 18. The right column (a) shows temperature from pressure-level ERA5 data (top), temperature from the ERA5-based LES with 6hour nudging (middle) and temperature from the ERA5-based LES with no nudging (bottom). The left column (b) shows profiles of temperature (left) and LWC (right) for different nudging timescales (τs).

figure1.png.

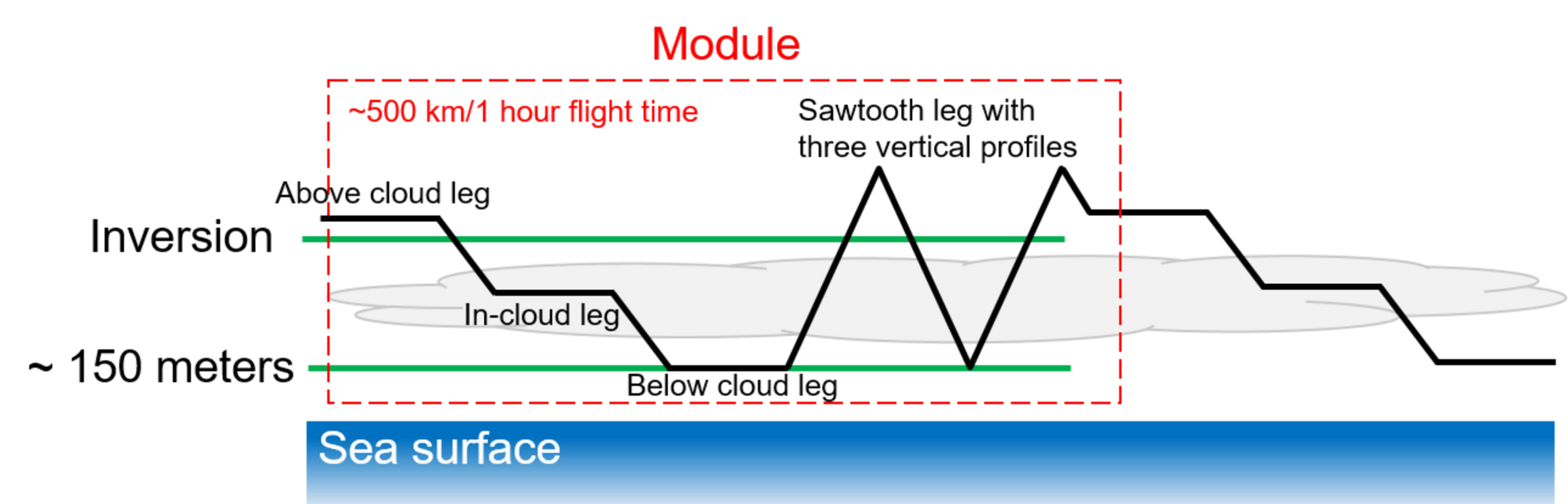

Module
~500 km/1 hour flight time
Sawtooth leg with
three vertical profiles
Above cloud leg
Inversion
In-cloud leg
~ 150 meters
Below cloud leg
Sea surface

figure2.jpg.

a)
b)
RSTB (SST proxy) (°C)
ERA5 reanalysis SST (°C)
Ship SST (°C)
ERA5 reanalysis SST (°C)

figure3.jpg.

Cloud Morphology Classification Decision Tree

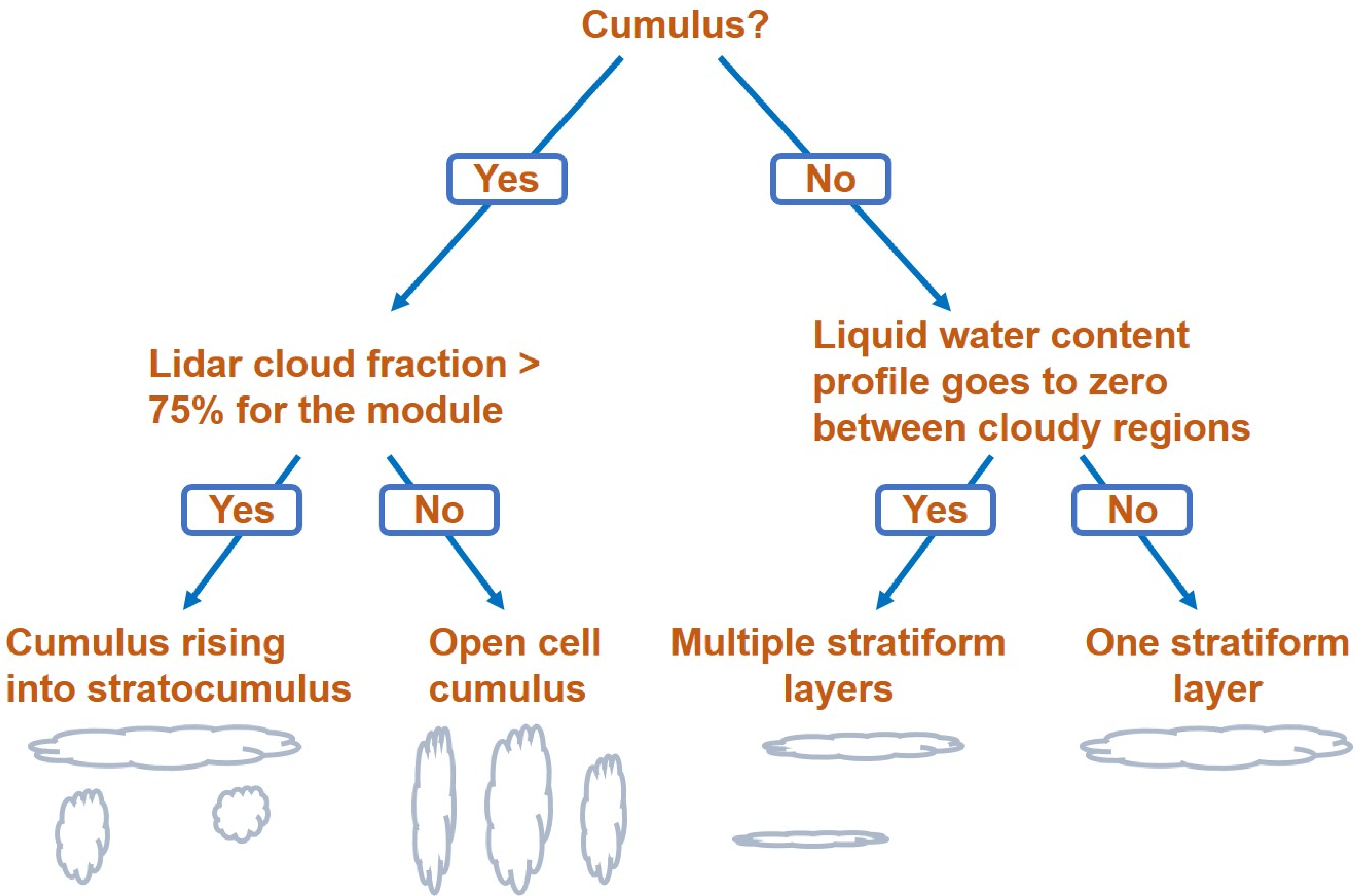


figure4.jpg.

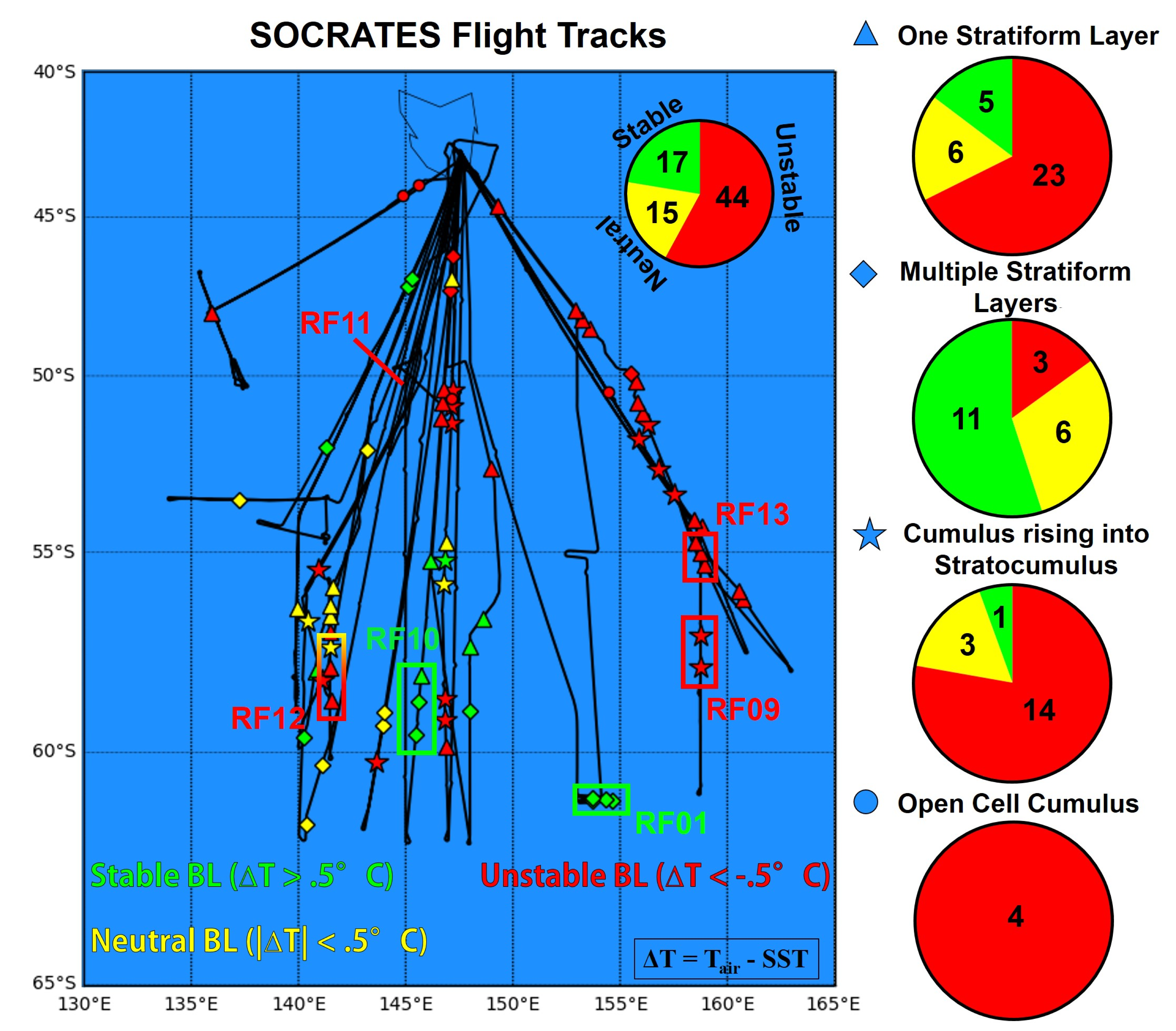

SOCRATES Flight Tracks
40°S
45°S
50°S
55°S
60°S
65°S
130°E
135°E
140°E
145°E
150°E
155°E
160°E
165°E
Stable
Unstable
Neutral
17
15
44
RF11
RF13
RF10
RF12
RF09
RF01
Stable BL (ΔT > .5° C)
Unstable BL (ΔT < -.5° C)
Neutral BL (|ΔT| < .5° C)
ΔT = T_air - SST
One Stratiform Layer
5
6
23
Multiple Stratiform Layers
3
11
6
Cumulus rising into Stratocumulus
1
3
14
Open Cell Cumulus
4

figure5.jpg.

RF01: Two layer stratus in a stable BL
RF09: Cumulus rising into stratocumulus in an unstable BL
RF10: Two layer stratus in a stable BL
RF11: Open cell cumulus in an unstable BL
RF12: Stratocumulus in an unstable/neutral BL
RF13: Stratocumulus in an unstable BL
Visible reflectance with sea level pressure
a)
b)
c)
d)
e)
f)
50°S
55°S
60°S
130°E
135°E
140°E
145°E
150°E
155°E
160°E
165°E
Aircraft Images
ERA5-based
3.2 km
12.8 km
12.8 km
Obs-based
No Obs-based LES for RF11

figure6.jpg.

Reflectivity from HIAPER Cloud Radar
Simulated reflectivity from ERA5-based LES
Simulated reflectivity from Obs-based LES
dBZ
50
17
8
−1
−10
−19
−28
−37
−60
a)
RF01: Two layer stratus in a stable BL
Z (km)
0
2
2.0
2.2
2.4
2.6
Hour of 01-16 2018 (UTC)
b)
RF09: Cumulus rising into stratocumulus in an unstable BL
2.8
3.0
3.2
3.4
3.6
3.8
4.0
Hour of 02-05 2018 (UTC)
c)
RF10: Two layer stratus in a stable BL
0.8
0.9
1.0
1.1
1.2
1.3
1.4
Hour of 02-08 2018 (UTC)
d)
RF11: Open cell cumulus in an unstable BL
3.00
3.25
3.50
3.75
4.00
4.25
4.50
4.75
5.00
Hour of 02-17 2018 (UTC)
No Obs-based LES for RF11
e)
RF12: Stratocumulus in an unstable/neutral BL
3.5
3.6
3.7
3.8
3.9
4.0
4.1
Hour of 02-18 2018 (UTC)
f)
RF13: Stratocumulus in an unstable BL
2.9
3.0
3.1
3.2
3.3
3.4
Hour of 02-20 2018 (UTC)

figure7.jpg.

RF12: Stratocumulus in an unstable/neutral boundary layer
RF13: Stratocumulus in an unstable boundary layer
a)
b)
c)
d)
Pressure (mb)
T (°C)
QV (g kg^{-1})
Cloud Fraction (-)
LWC (g m^{-3})
Obs
ERA5-based LES
Obs-based LES
ERA5
CAM6
AM4

figure8.jpg.

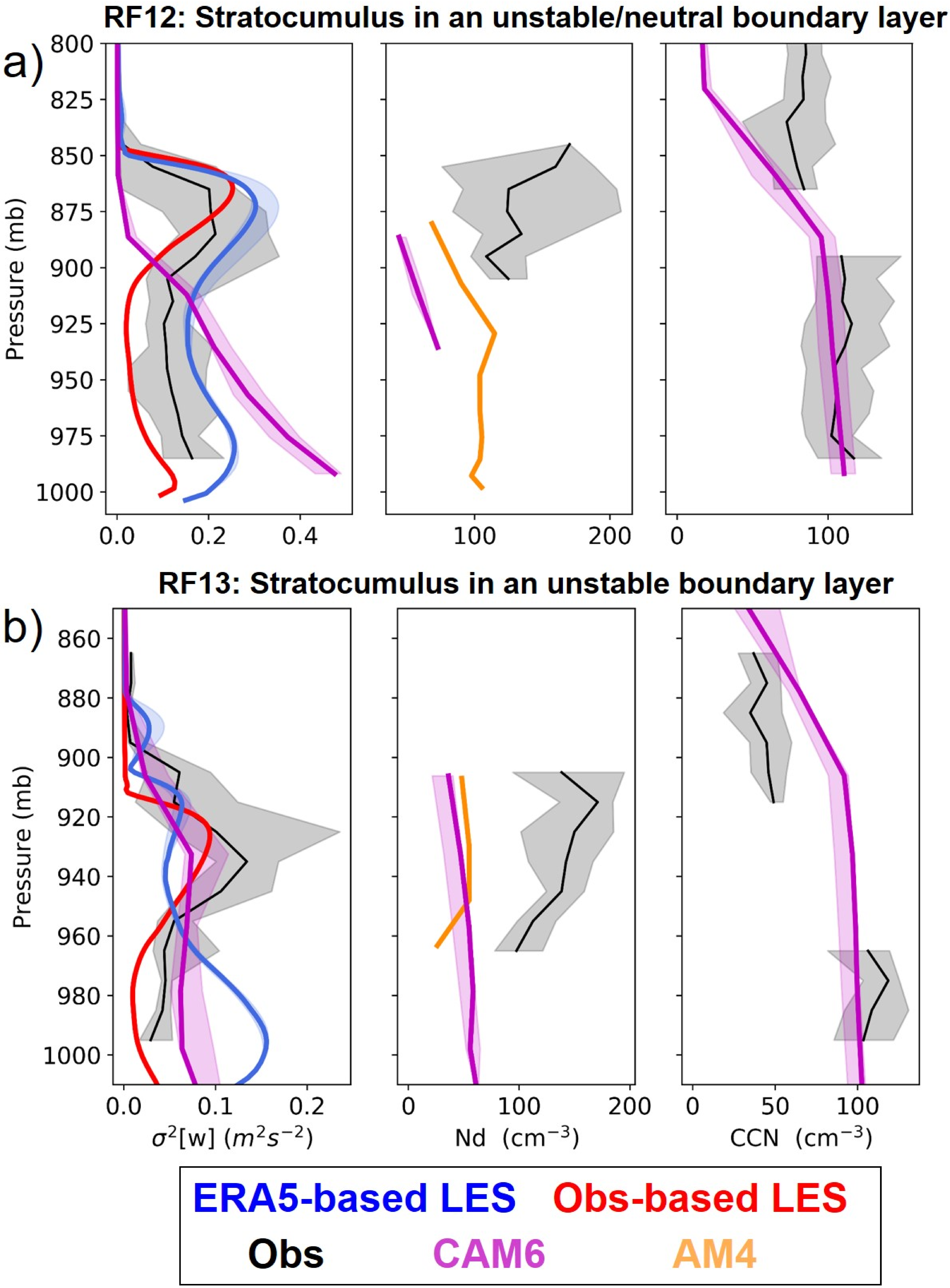

RF12: Stratocumulus in an unstable/neutral boundary layer
a)
Pressure (mb)
800
825
850
875
900
925
950
975
1000
0.0
0.2
0.4
100
200
0
100
RF13: Stratocumulus in an unstable boundary layer
b)
Pressure (mb)
860
880
900
920
940
960
980
1000
0.0
0.1
0.2
0
100
200
0
50
100
σ^2[w] (m^2s^{-2})
Nd (cm^{-3})
CCN (cm^{-3})
ERA5-based LES
Obs-based LES
Obs
CAM6
AM4

figure9.jpg.

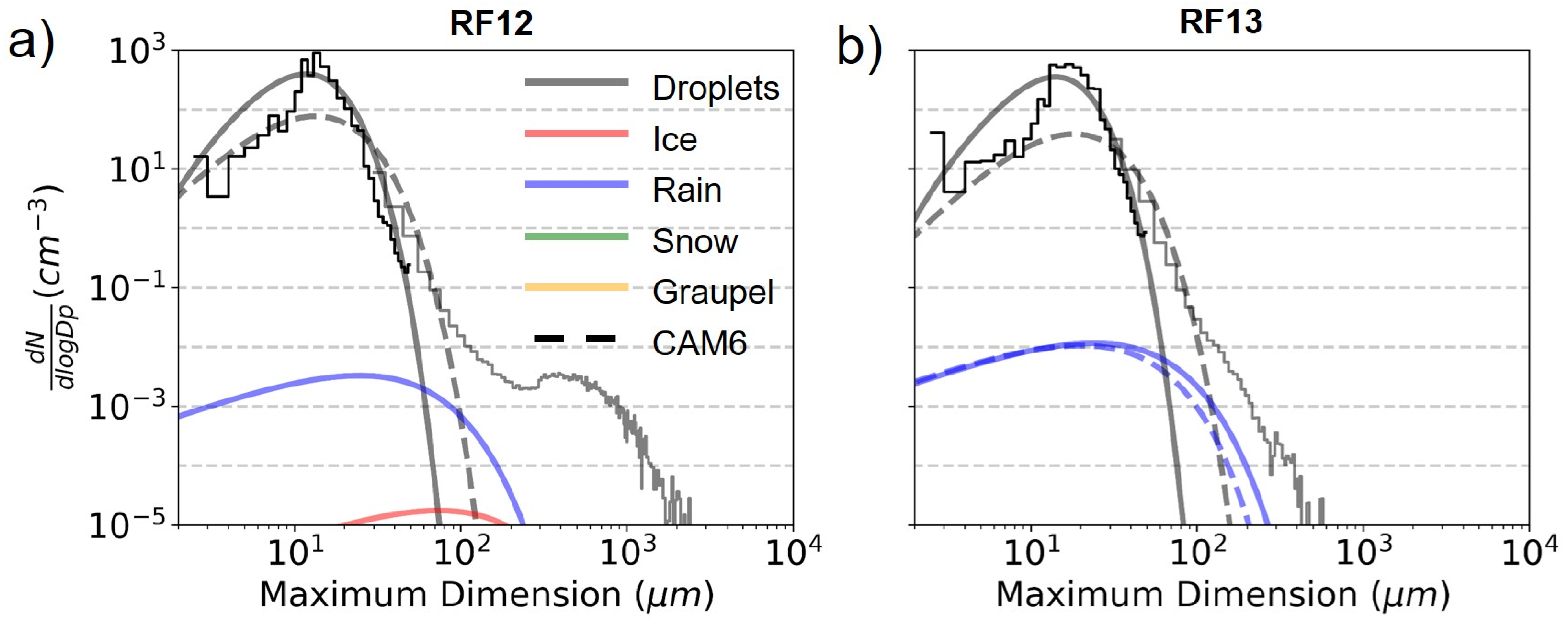

a)
RF12
b)
RF13
Droplets
Ice
Rain
Snow
Graupel
CAM6
$\frac{dN}{dlogDp}(cm^{-3})$
10^{3}
10^{1}
10^{-1}
10^{-3}
10^{-5}
10^{1}
10^{2}
10^{3}
10^{4}
Maximum Dimension (μm)
Maximum Dimension (μm)

figure10.jpg.

RF01: Two layer stratus in a stable boundary layer
RF10: Two layer stratus in a stable boundary layer
a)
b)
c)
d)
Pressure (mb)
T (°C)
QV (g kg^{-1})
Cloud Fraction (-)
LWC (g m^{-3})
Obs
ERA5-based LES
Obs-based LES
ERA5
CAM6
AM4

figure11.jpg.

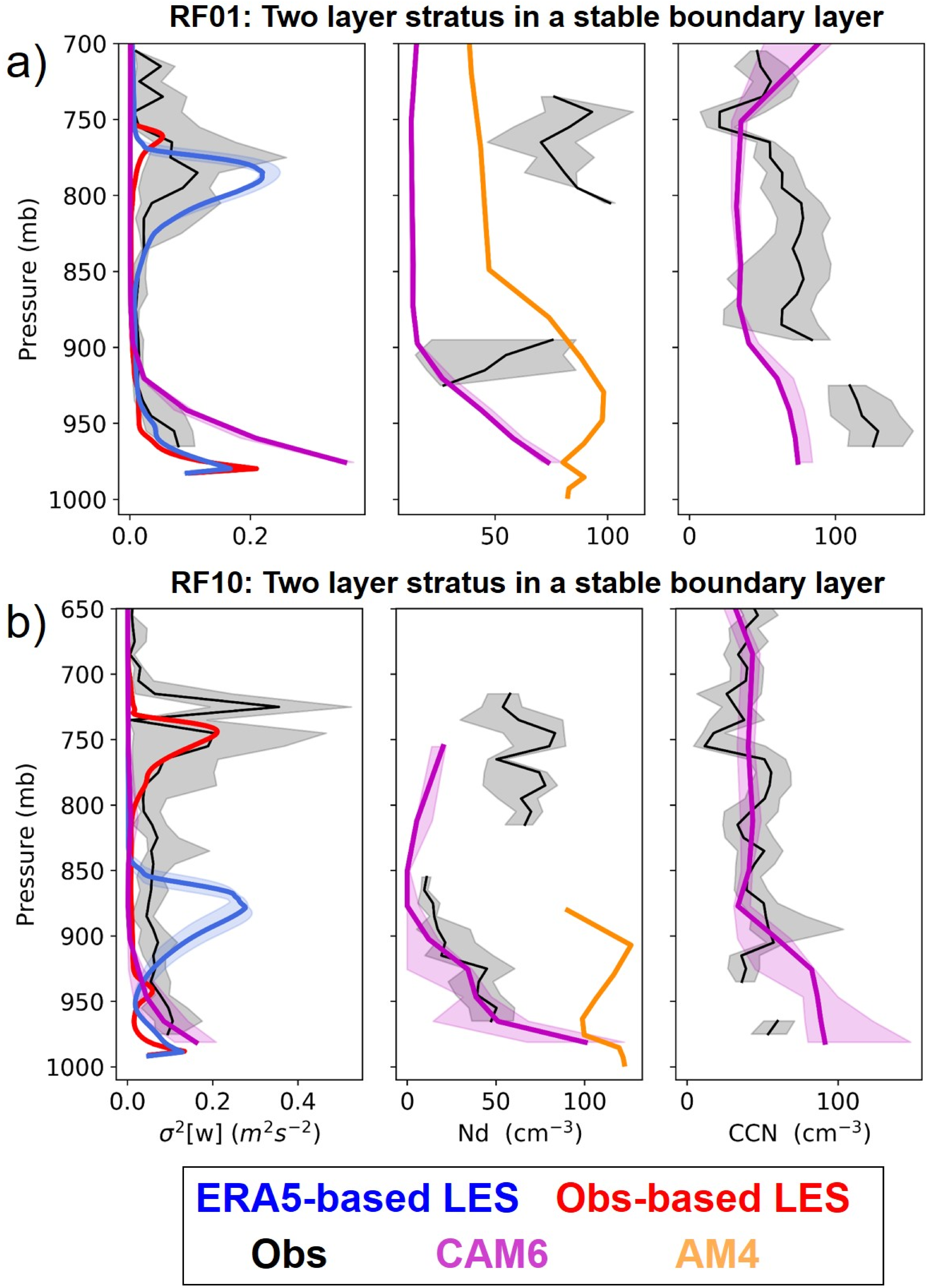
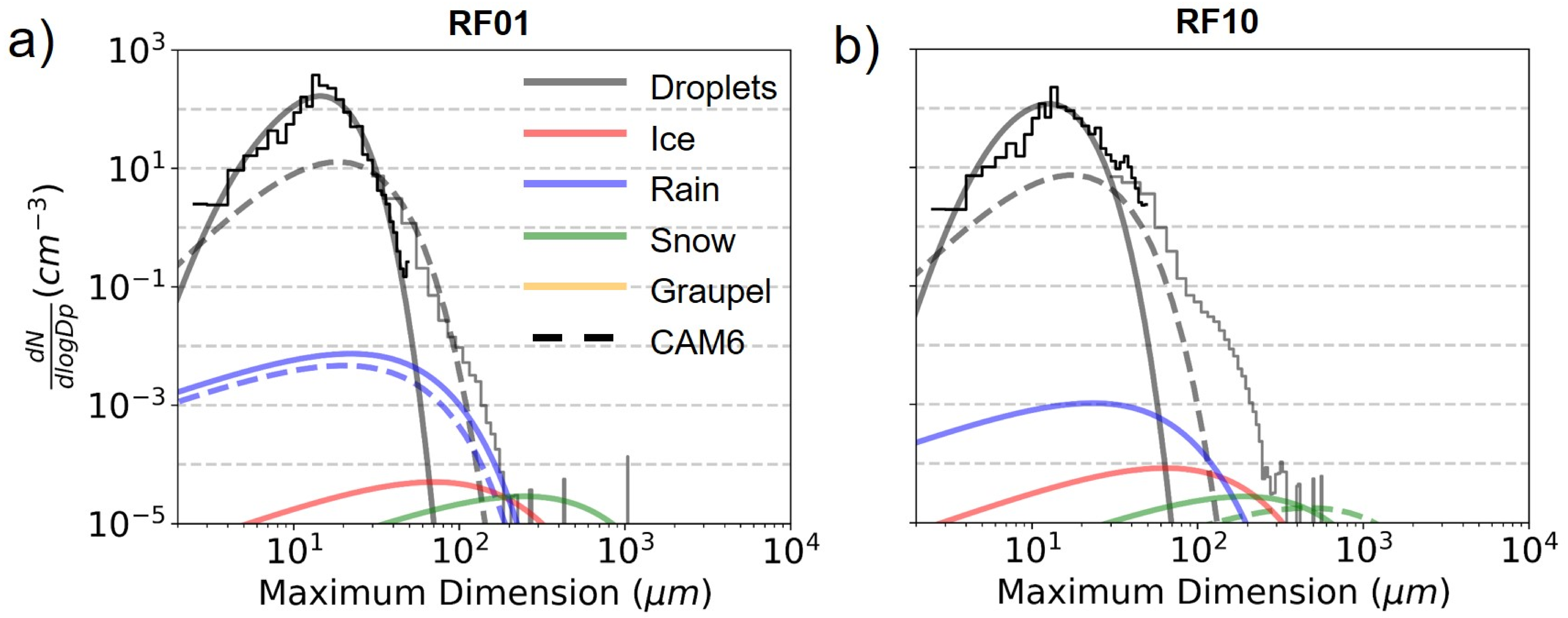

RF01: Two layer stratus in a stable boundary layer
a)
Pressure (mb)
700
750
800
850
900
950
1000
0.0
0.2
50
100
0
100
RF10: Two layer stratus in a stable boundary layer
b)
Pressure (mb)
650
700
750
800
850
900
950
1000
0.0
0.2
0.4
0
50
100
0
100
σ²[w] (m^2s^{-2})
Nd (cm⁻³)
CCN (cm⁻³)
ERA5-based LES
Obs-based LES
Obs
CAM6
AM4

figure12.jpg.



a)
RF01
b)
RF10
Droplets
Ice
Rain
Snow
Graupel
CAM6
$\frac{dN}{dlogDp}(cm^{-3})$
10^{3}
10^{1}
10^{-1}
10^{-3}
10^{-5}
10^{1}
10^{2}
10^{3}
10^{4}
Maximum Dimension (μm)
Maximum Dimension (μm)

figure13.jpg.

RF09: Cumulus rising into stratocumulus in an unstable boundary layer
RF11: Open cell cumulus in an unstable boundary layer
a)
b)
c)
d)
Pressure (mb)
T (°C)
QV (g kg⁻¹)
Cloud Fraction (-)
LWC (g m⁻³)
Obs
ERA5-based LES
Obs-based LES
ERA5
CAM6
AM4

figure14.jpg.

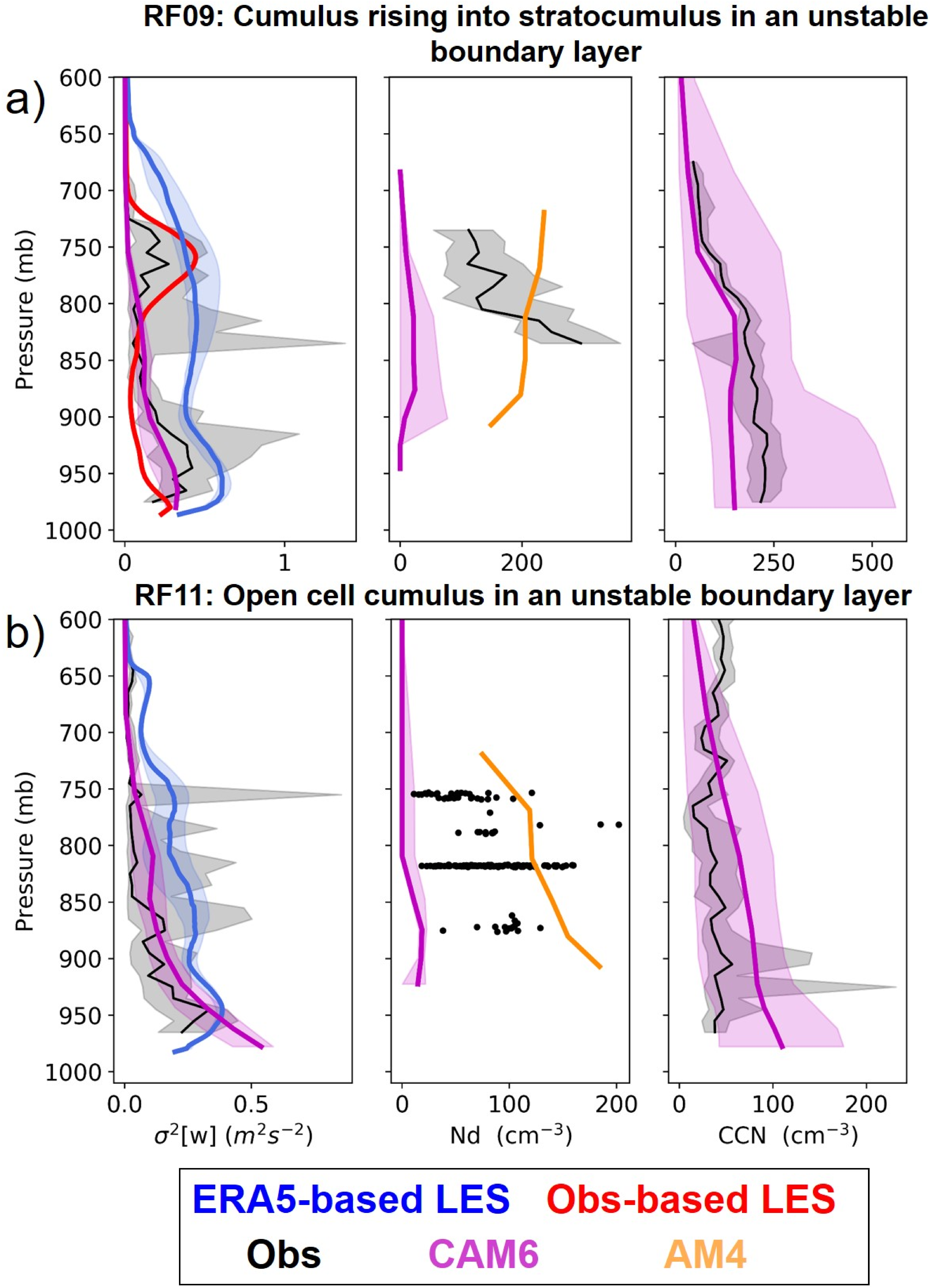
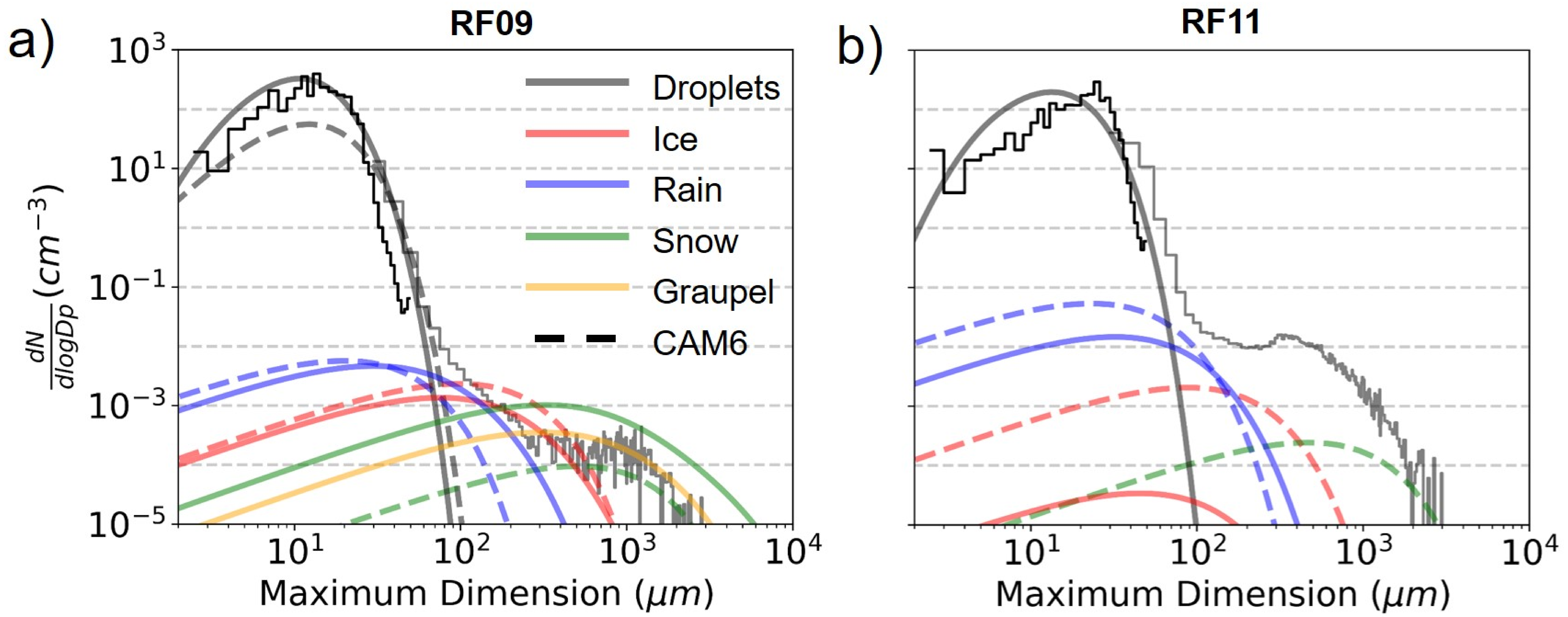

RF09: Cumulus rising into stratocumulus in an unstable boundary layer
a)
Pressure (mb)
600
650
700
750
800
850
900
950
1000
0
1
0
200
0
250
500
RF11: Open cell cumulus in an unstable boundary layer
b)
Pressure (mb)
0.0
0.5
0
100
200
0
100
200
σ^2[w] (m^2s^{-2})
Nd (cm^{-3})
CCN (cm^{-3})
ERA5-based LES
Obs-based LES
Obs
CAM6
AM4

figure15.jpg.

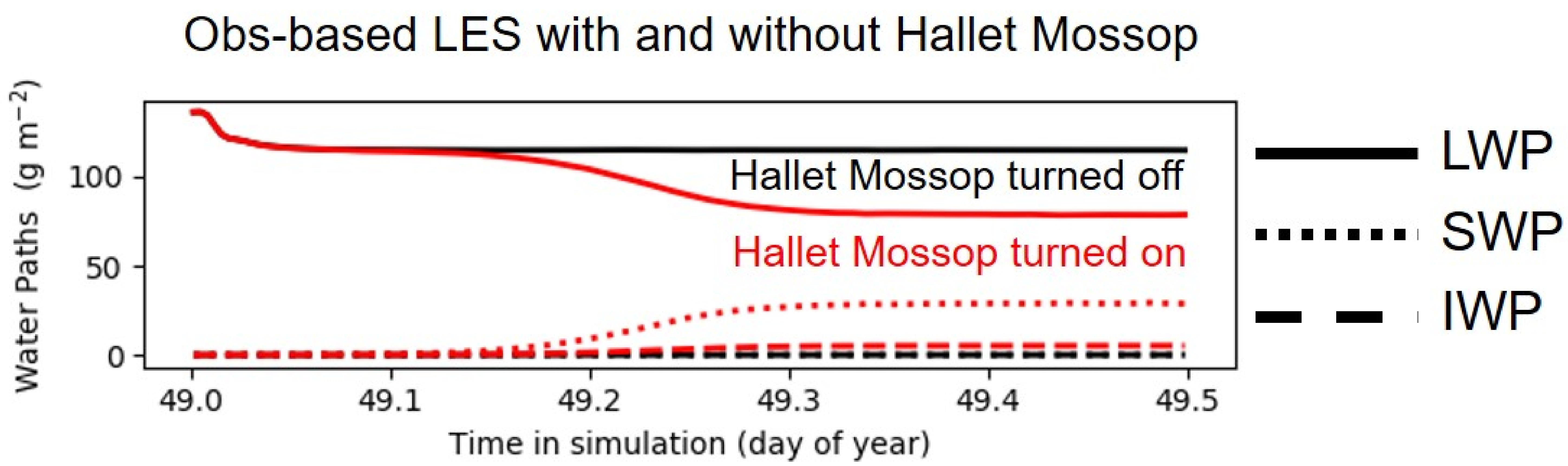


a)
RF09
b)
RF11
Droplets
Ice
Rain
Snow
Graupel
CAM6
$\frac{dN}{dlogDp}(cm^{-3})$
10^{3}
10^{1}
10^{-1}
10^{-3}
10^{-5}
10^{1}
10^{2}
10^{3}
10^{4}
Maximum Dimension (μm)
Maximum Dimension (μm)

figure16.jpg.

Microphysics Sensitivity Tests with the Obs-based LES for case RF12

figure17.jpg.



Obs-based LES with and without Hallet Mossop
Water Paths (g m⁻²)
100
50
0
Hallet Mossop turned off
Hallet Mossop turned on
49.0
49.1
49.2
49.3
49.4
49.5
Time in simulation (day of year)
LWP
SWP
IWP

figure18.jpg.

a) Input and output temperature from the LES
ERA5 reanalysis data
ERA5-based LES with 6hour nudging
ERA5-based LES with no nudging
Pressure (mb)
Temperature (°C)
Hour of February 18th, 2018 (UTC)
b) Biases in mid-tropospheric temperature and LWC
Pressure (mb)
T (°C)
LWC (g kg−1)
Obs
No nudging
24hour Nudging
12hour Nudging
6hour Nudging