JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2020MS002205

Key Points:
- SAM LES represents diverse Southern Ocean boundary layer and cloud structures well
- CAM6 and AM4 maintain supercooled water in stratiform clouds but excessively glaciate cumuli
- CAM6 underpredicts stratiform cloud-driven turbulence and cloud droplet concentration

Correspondence to:
R. L. Atlas,
ratlas@uw.edu

Citation:
Atlas, R. L., Bretherton, C. S., Blossey, P. N., Gettelman, A., Bardeen, C., Lin, P., & Ming, Y. (2020). How well do large-eddy simulations and global climate models represent observed boundary layer structures and low clouds over the summertime Southern Ocean? *Journal of Advances in Modeling Earth Systems*, *12*, e2020MS002205. https://doi.org/10.1029/2020MS002205

Received 15 JUN 2020
Accepted 12 OCT 2020
Accepted article online 15 OCT 2020

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], and Yi Ming[4,5]

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

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 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 (fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis, "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 are 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 multilayer 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 tweak the Hallett-Mossop parameterization so that it activates within boundary layer clouds, and we achieve better agreement between observed and simulated microphysics. The nudged global climate models (GCMs) simulate liquid-dominated mixed-phase clouds in the stratiform cases but excessively glaciate cumulus clouds. Both GCMs struggle to represent two-layer clouds, and CAM6 has low droplet concentrations in all cases and underpredicts 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 and Aerosol Transport Experimental Study (SOCRATES) to evaluate two leading U.S. global climate models, the GFDL 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 multiplication process, called Hallett-Mossop rime splintering, is not active enough. CAM6 and AM4 make too much ice in, 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 45–65°S latitude band. They increase the albedo of the Earth, reduce sea surface temperatures (SSTs), and moderate global oceanic heat uptake (Hyder et al., 2018; Roemmich et al., 2015). 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 (Naud et al., 2014; Trenberth & Fasullo, 2010).

Insufficient Southern Ocean cloudiness in GCMs has been attributed to the lack of supercooled 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). This 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 1,000 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 (DeMott et al., 2016; McCluskey et al., 2018).

Southern Ocean boundary layer clouds are commonly mixed-phase, containing supercooled 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 supercooled 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 increased the amount of supercooled water within Southern Ocean clouds and the climate sensitivity. 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 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, which overlapped during January–February 2018 (McFarquhar et al., 2020). Here, we use a unique multisensor suite of SOCRATES observations to build case studies in different types of Southern Ocean cloudy boundary layers and compare them with two GCMs, Geophysical Fluid Dynamics Laboratory (GFDL) Atmosphere Model Version 4 (AM4) and National Center for Atmospheric Research (NCAR) Community Atmosphere Model Version 6 (CAM6), run in a nudged-meteorology mode and sampled at the locations and times of the airborne sampling.

One 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) is a complementary modeling strategy. LES 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. Thus, we also compare a suitably forced LES with the GCMs and aircraft observations. We identify strengths and

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 (SST proxy) [RSTB]
Microphysical properties	Ultra-high sensitivity aerosol spectrometer (UHSAS)	Large aerosol concentration/CCN [CONCU100_LWII]
	Cloud Droplet Probe (CDP)	Size distribution (2–50 μm) [CCDP_RWIO]
	Two-Dimensional Stereo Probe (2-DS)	Size distribution (25 μm to 1 mm)
Dynamics	Pressure ports, inertial reference systems and GPS	Horizontal and vertical wind [UIC,VIC,WIC]

Note. Variable Names From NCAR EOL Aircraft Data Files Are Included in Square Brackets in the Right Column.

weaknesses of 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 part 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 supercooled water in Southern Ocean mixed-phase clouds more realistically than CAM5 and reproduces the shape of SOCRATES droplet and ice crystal size distributions but with lower number concentrations than observed. Here, we use LES to provide process-level explanations for some of the successes and failures of CAM6 and AM4 discussed in Z2020 and G2020, especially pertaining to boundary layer structure, mixed-phase microphysics, and turbulence.

2. Observations

During SOCRATES, the U.S. National Science Foundation Gulfstream-V (G-V) research aircraft, operated by the Research Aviation Facility of NCAR, was based in Hobart, Tasmania, at 43°S, 145°E. The G-V conducted 120 hr of in situ sampling below, in, and above diverse cold sector Southern Ocean clouds between 45°S and 62°S during 15 January to 25 February 2018. The aircraft instrumentation and flight plans were targeted for 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, and a horizontal resolution of 120–180 m, depending on aircraft ground speed. The SOCRATES instruments and measurements used in this study are listed in Table 1.

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 10 min level legs—150 m above the cloud top (above-cloud leg), within the cloud layer (in-cloud leg), and 150 m 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 CAPRICORN-2 campaign.

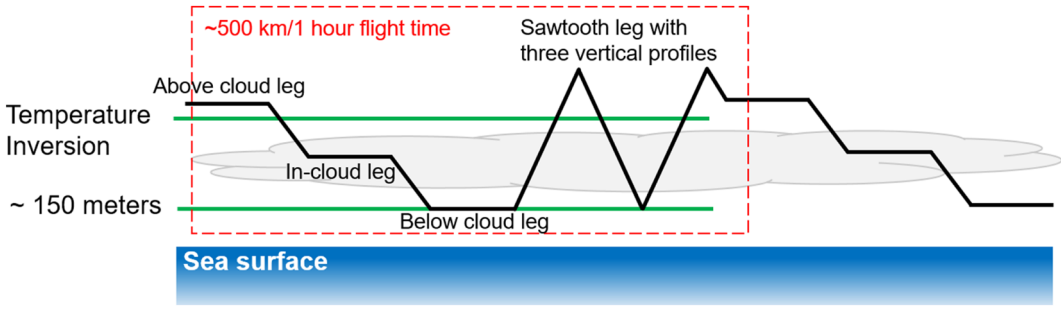

Figure 1. Ideal SOCRATES flight module (dashed red box) with a below-cloud leg, in-cloud leg, above-cloud leg, and sawtooth leg comprised of profiles that extend from the subcloud layer to the free troposphere.

2.2. Vertical Wind 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 at least 0.1 m s^{-1} (Cooper et al., 2016), comparable to typical vertical wind speeds, but variability in the vertical wind is more accurately measured.

Traditionally, vertical wind variance has been estimated from aircraft data using long level legs through relatively homogeneous turbulence. The SOCRATES flights involved extensive profiling, and the boundary layer clouds often had substantial mesoscale variability and large-scale gradients, so we developed a modified estimation method. We computed the running variance in w over a 20 s block centered around the measurement time ($\sigma^2_{20}[w]$), 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 300m min^{-1}. During a 20 s block, its altitude changed by 100 m, so $\sigma^2_{20}[w]$ encompasses an altitude range much narrower than the typical depth of boundary layers observed during SOCRATES (>1 km). We correct $\sigma^2_{20}[w]$ to account for the power contained within periods larger than 20 s, as described in Appendix A. We refer to the corrected vertical wind variance, which is an estimate of the total vertical wind variance, as $\widetilde{\sigma^2[w]}$.

2.3. ERA5 Reanalysis and Its Application

We use the fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis of the global climate (ERA5) for the SOCRATES period to evaluate aircraft measurements and to initialize and force our LES cases, as detailed in section 4.2. 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.

The G-V radiometric surface temperature (RSTB) is 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 positive temperature-dependent bias in the RSTB which approaches 2°C at the lowest 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 data set. In contrast, the ERA5 SST is unbiased compared with coincident measurements from the *R/V Investigator* during CAPRICORN-2 (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 SOCRATES 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.

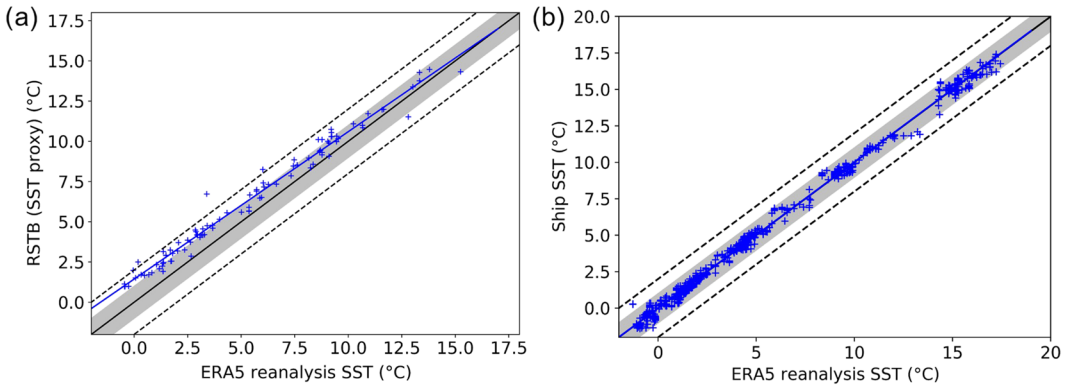

Figure 2. ERA5 SST scattered with (a) radiometric surface temperature (RSTB) from the aircraft and (b) SST from the *R/V Investigator*. Blue crosses are the data; the blue line is the best fit line, and the black line is the 1-to-1 line. Within the shaded region, the difference between the SST estimates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

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 were both commonly observed during SOCRATES 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 m altitude up past the temperature inversion base. For profiles with multiple temperature inversions, the aircraft must reach up past the highest cloud layer subjectively included within the boundary layer. We estimate the air-sea temperature difference by subtracting the ERA5 SST from the ERA5 2 m air 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.

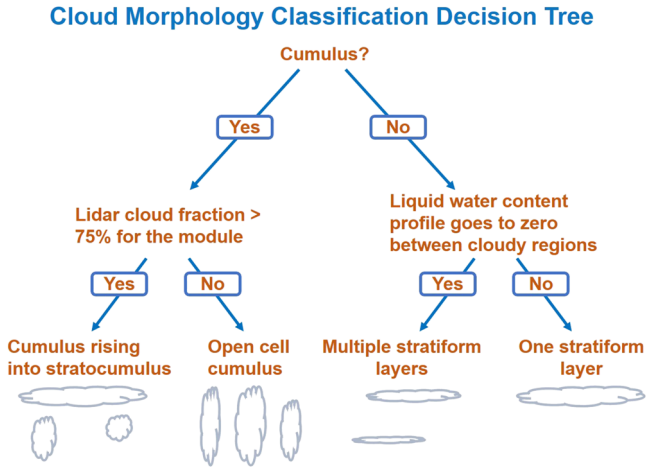

Figure 3. Decision tree showing how aircraft measurements of LWC, vertical wind, vertical wind variance, and lidar data from the HSRL are used to classify the cloud morphology sampled in every vertical profile from SOCRATES, into one of four categories.

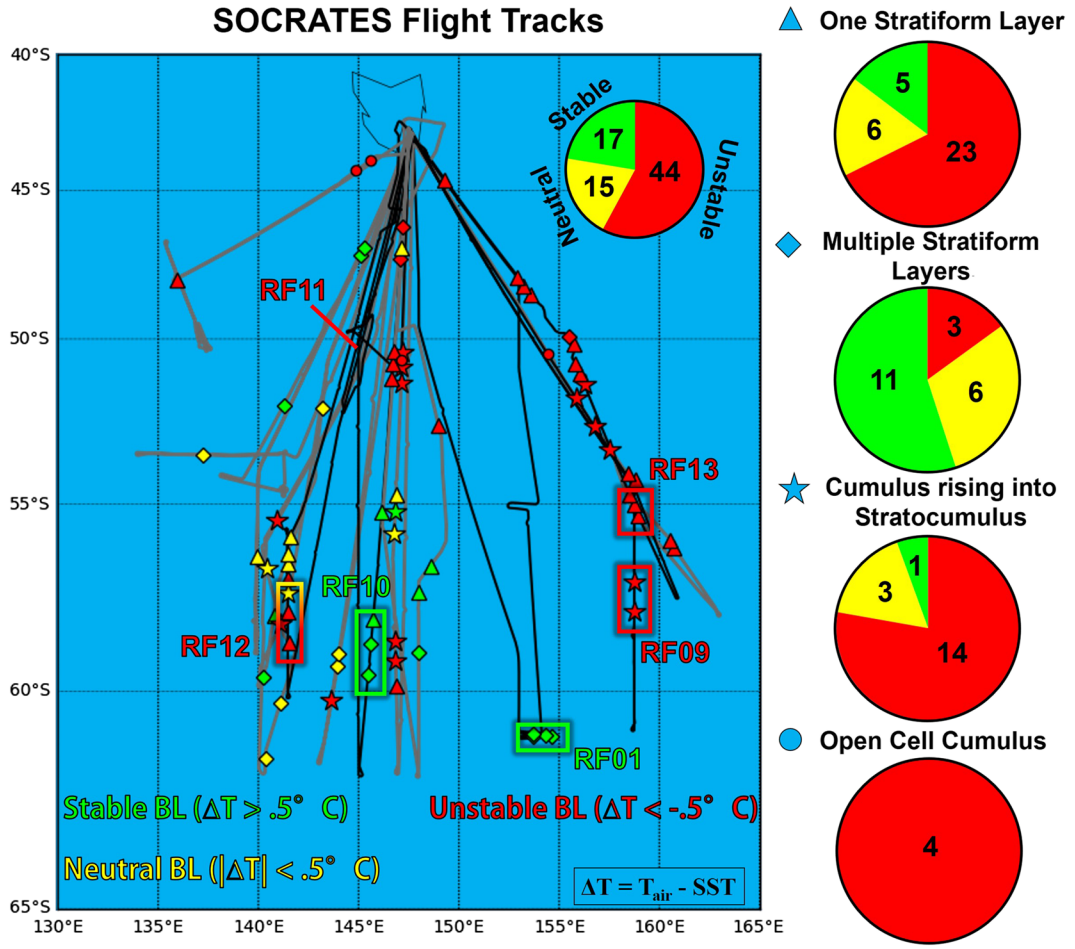

Figure 4. Map of the Southern Ocean between Antarctica and Tasmania with SOCRATES flight tracks (black and gray lines) and symbols showing the location of the vertical aircraft profiles from SOCRATES that profiled the entire boundary layer and sampled cloud. Black lines show the flight tracks for the flights that have been developed into case studies. 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.

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 s running vertical wind variance ($\widehat{\sigma^2[w]}$) into 2 mb pressure bins that span the range from 1,050 to 400 mb. This binning substantially reduces the noise in the measurements while still resolving sharp temperature inversions. We calculate the bin medians $\overline{LWC}$, $\bar{w}$, and $\overline{\widehat{\sigma^2[w]}}$. If any pressure bins simultaneously have cloud ($\overline{LWC} > 0.01$ g kg^{-1}), a strong updraft ($\bar{w}$>1 m s^{-1}) (LeMone & Zipser, 1980), and turbulence ($\overline{\widehat{\sigma^2[w]}}$> 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 an 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 without cloud situated between bins containing cloud, 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 data set,

Table 2
Summary of SOCRATES Cases

Flight	Date/Time (2018/UTC)	Reference Time	Location	Stability	Cloud Regime	T_{top} (°C)	N_d (cm^{-3})	U_{950} ($m\ s^{-1}$)	ΔT (°C)
RF01	16 Jan 1:50–2:45	2:30	61°S, 154.25°E	Stable	Two St Layers	−3.3/−11.8	75	19	1.44
RF09	5 Feb 2:40–4:00	3:30	58°S, 158.75°E	Unstable	Cu under Sc	−18.2	190	15	−4.05
RF10	8 Feb 0:00–1:25	1:00	59.5°S, 145.5°E	Stable	Two St Layers	0.6/−13.4	55	17	1.87
RF11	17 Feb 3:00–5:00	4:00	51°S, 144.5°E	Unstable	Open-Cell Cu	N/A	115	17	−1.48
RF12	18 Feb 3:25–4:10	4:00	57°S, 141.5°E	Unstable/Neutral	Sc	−7.5	210	17	−0.70
RF13	20 Feb 2:50–3:30	3:00	54.75°S, 158.5°E	Unstable	Sc	−1.4	180	8	−0.95

Note. T_{top} = cloud top temperature, N_d = median number concentration, U_{950} = horizontal wind speed at 950 mb, ΔT = air-sea temperature difference (SST − surface air temperature).

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 low 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 occasionally occur in neutral and stable boundary layers during SOCRATES, when 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 labeled 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 two to four vertical profiles which observed similar boundary layer and cloud properties throughout the sampling period. Since the selected cases feature similar conditions over hundreds of kilometers, it is meaningful to compare them with the coarse resolution ERA5 reanalysis and the nudged GCMs. 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 (Schmidt et al., 2017). Cloud top temperatures (T_{top}) of the uppermost sampled cloud layers range from −1°C to −18°C, and air-sea temperature differences (ΔT) range from −4°C (unstable) to 2°C (stable). All cases feature supercooled water-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

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 CAM6 and AM4. 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 the UM5 advection scheme (Yamaguchi et al., 2011), RRTMG radiation (Mlawer et al., 1997), and Morrison two-moment microphysics with graupel (Morrison et al., 2005) (hereafter M2005). We use a model time step of 5 s, a radiation time step of 15 s, and a uniform SST specified from ERA5.

The microphysics scheme in M2005 includes a parameterization of Hallett-Mossop rime splintering (Hallett & Mossop, 1974) relevant to some SOCRATES mixed-phase clouds. This scheme allows new ice particles to splinter from rimed particles 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 can occur on both snow and graupel if they have hydrometeor masses larger than 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. We perform sensitivity tests by removing these three thresholds in section 6.

Simulations are run for 12–14 hr, long enough to spin up realistic turbulence and boundary layer mixing but not so long as to drift away from the rapidly evolving observed boundary layer. We use a horizontal resolution of 50 m and a doubly periodic square domain of length 12.8 km for all cases. This relatively small domain size was chosen for computational efficiency. It is too small to accurately represent internally generated mesoscale cellular patterns of scale 20–100 km that are often visible in Southern Ocean clouds, but those patterns take longer than the simulated time of 12 hr to fully develop (Zhou & Bretherton, 2019).

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). Near the surface, we choose a vertical grid spacing of 25 m, within a factor of 2 of the horizontal resolution, to better represent near-surface isotropic turbulence that efficiently transfers heat and moisture to the rest of the boundary layer (Berner et al., 2015).

The domain height is at least twice the height of the boundary layer to provide an overlying layer for gravity wave damping. Two cases with shallow cloud-topped boundary layers, RF12 and RF13, are run on a 192-level vertical grid. The other four cases have clouds extending through deeper boundary layers; they are run on a 320-level vertical grid.

4.2. LES Initialization and Forcing

To account for the thermodynamics of vapor-liquid phase change, SAM uses saturation adjustment and prognoses two moist-conserved variables. The first is 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. The second is liquid water temperature ($T_L = T - Lq_c/c_p$). Precipitation-sized particles are separately prognosed and may be rain, snow, or graupel.

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 quickly drift away from reality. After considerable experimentation, we used two LES forcing methodologies, with complementary advantages, for each case.

An ERA5-based simulation is initialized and forced exclusively with ERA5 reanalysis data and run for 14 hr, reaching the reference time at hour 12. This is an analog 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, humidity, and cloud liquid water match those of a single aircraft profile, which we refer to as the reference profile. The LES q_t and T_L are initialized from a lightly smoothed version of this profile (see Appendix B for details) and then their horizontal domain means are aggressively nudged toward it with a 20 min timescale. The Obs-based LES allows comparison of the simulated and observed microphysics without having to account for major differences in the cloud macrophysics. The solar zenith angle is held constant at the reference time of the case. 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.

We specify a uniform cloud droplet number concentration for each case, given in Table 2, which is equal to the median droplet number concentration of the reference profile. For the shallow cumulus case RF11, we instead use the median droplet number concentration from all of the in situ data from the flight.

Horizontal winds, surface pressure, and SST from ERA5 are used in both the Obs-based and ERA5-based simulations. Vertical wind is derived from the ERA5 pressure velocity as described in Appendix C. Both LES experiments are forced by ERA5-derived geostrophic winds and nudged toward the ERA5 horizontal winds with a nudging timescale of 1 hr for the ERA5-based simulation and 20 min for the Obs-based simulation.

All ERA5 data were obtained on pressure levels and averaged over a 1° box (including 16 ERA5 columns) centered on the model domain.

4.3. Description of CAM6 and AM4 GCMs

Our other goal of this study is to evaluate the atmospheric components of two GCMs, CAM6 (Danabasoglu et al., 2020) and AM4 (Zhao et al., 2018), that have been run in hindcast mode for the SOCRATES experiment and lightly nudged to ERA5 reanalysis. Both models use a finite volume dynamical core and comparable grid resolutions.

CAM6 is run on a $0.9° \times 1.25°$ 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 (Morrison & Gettelman, 2008) (hereafter MG2008) is analogous to the M2005 scheme in SAM but optimized for a GCM framework. Unlike M2005, MG2008 does not include graupel. However, M2005 does not produce substantial concentrations of graupel in any of the SOCRATES cases (now shown). In contrast to the case-specified droplet concentration used for the LES, CAM6 predicts aerosol using the Modal Aerosol Module Version 4 (MAM4) (Liu et al., 2016), initialized from climatological profiles in year 2000 from the Coupled Model Intercomparison Project phase 6 (CMIP6) emissions inventory, and explicitly activates cloud droplets. CAM6 is subsampled along the SOCRATES flight track such that for every 10 min 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 entraining detraining 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, ice water content, and cloud droplet number concentration. AM4 microphysics follows Rotstayn (1997) for hydrometeor mass (which is diagnostic) and cloud fraction and Ming et al. (2006, 2007) for droplet number 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 both GCMs, horizontal winds, air temperature, and SST are nudged with a 24 hr timescale to ERA5 reanalysis.

5. Model-Observation Comparisons

We compare how well the SAM LES and the two GCMs can represent observable physical processes important for the formation, evolution, and radiative properties of Southern Ocean boundary layer clouds across our set of cases. Section 5.1 qualitatively describes the synoptic environment in which the clouds have formed and evolved for each case, using ERA5 reanalysis and satellite imagery. Sections 5.2–5.6 then use SOCRATES in situ observations to evaluate boundary layer structure and turbulence, cloud macrophysics, and cloud microphysics in the models.

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. Three cases with stratocumulus-topped unstable boundary layers, RF09, RF12 and RF13, are in southwesterly flow with cold advection. Two other cases sampled boundary layers with more unstable air-sea temperature differences (Table 2). Of those cases, RF09 has cumulus rising into a stratocumulus layer, and RF11 features open cell cumulus in westerly flow nearer to Tasmania.

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 forcing 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 the horizontal mean to have the vertical profile of humidity,

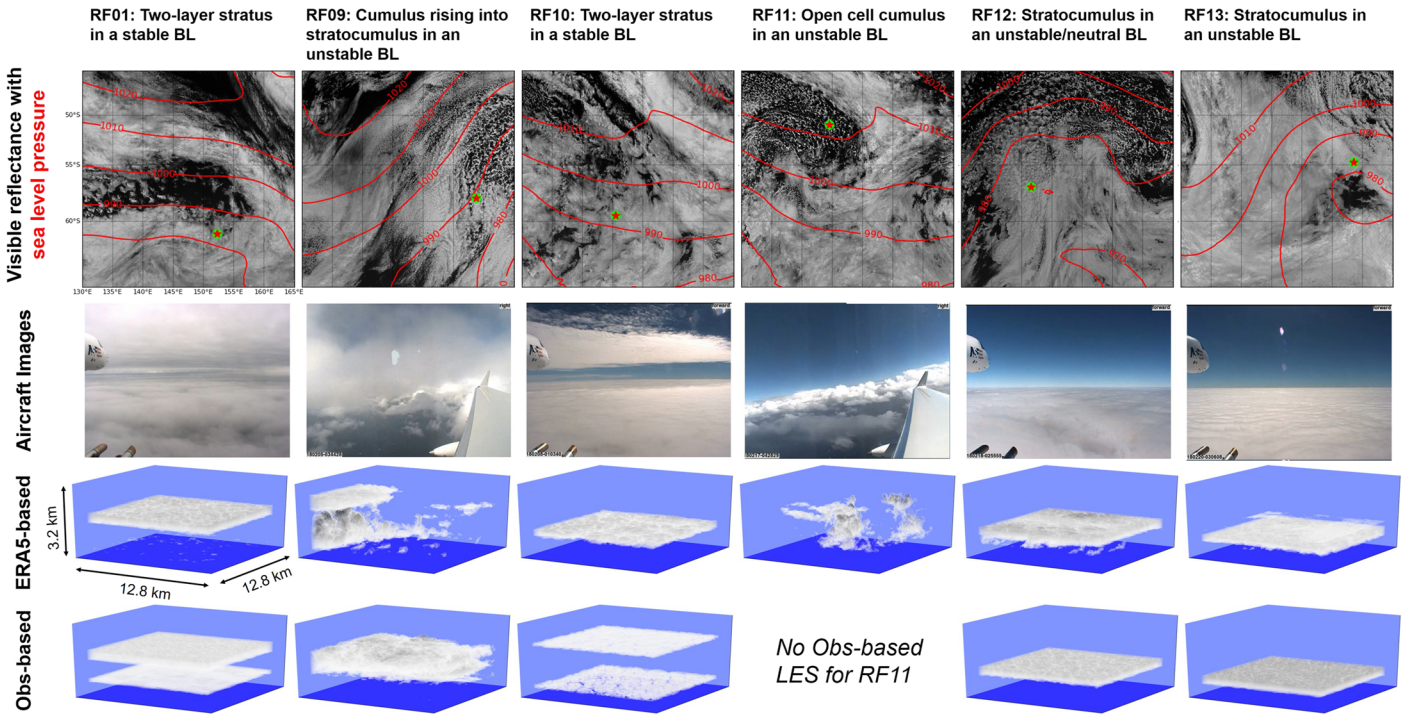

Figure 5. Top row: visible reflectance from Himawari with contours of sea level pressure from ERA5 reanalysis. The stars inscribed in the circles show the locations of the cases. Second row: snapshots during the flight module corresponding to each case from the aircraft's front facing, right side or left side camera. Third row: LWC at hour 12 of the ERA5-based simulations, which correspond to the reference times of each case. Fourth row: LWC at hour 12 of the Obs-based simulations, at which point all of the simulations have been in a steady state for at least 2 hr.

temperature, and cloud from an observed sounding. Hence, it consistently and accurately simulates the cloud morphology of observed stratiform cloud layers (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 to either temperature or humidity, 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 well represented by ERA5's relatively coarse vertical grid.

Figure 6 shows a comparison of simulated reflectivities from both LES experiments and observed reflectivities from the G-V cloud radar, for all six cases. The yellow line indicates the reference profile, 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 two-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, which differs considerably between the two LES experiments. 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. Z2020 used the COSP radar simulator in CAM6 to show that high reflectivities in SOCRATES clouds are due to large frozen particles. We will

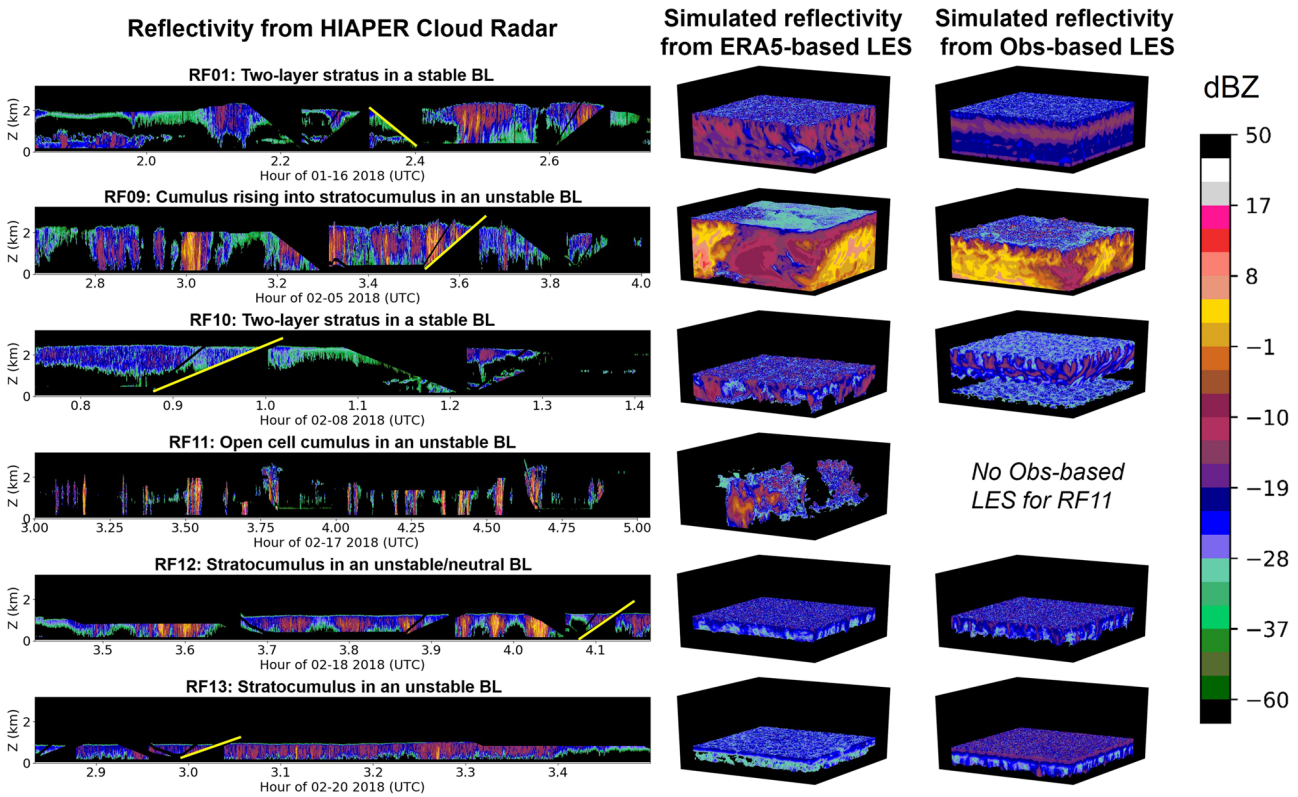

Figure 6. Observed reflectivities (left column) and simulated reflectivities from the ERA5 and Obs-based LES experiments (right columns) for all selected SOCRATES cases. Yellow lines indicate the reference profiles to which the Obs-based LES simulations are nudged. Small LES domains do not capture mesoscale variability and are best compared with the observed reflectivity profiles along the yellow lines.

show in sections 5.6 and 6 that there are too few large frozen particles in the LES in RF11 and RF12, due to Hallett-Mossop rime splintering being inactive.

5.2. Thermodynamic Structure of the Boundary Layer

We use SOCRATES observations to evaluate mean temperature and humidity profiles in the LES experiments, the GCMs, and ERA5 reanalysis, to which the models are nudged. The first and third columns within each panel in Figure 7 show the observed temperature and humidity reference profiles (solid black lines), respectively. The gray shaded areas show the 10th to 90th percentile of the aircraft observations from the surrounding module, calculated using 10 mb pressure bins. The reference profile is an interpolation between a single aircraft profile and ERA5 reanalysis so it is sometimes outside of the 10th to 90th percentile range of the module observations. Since there is no reference profile for RF11, the solid black line is instead the median of the aircraft observations. The yellow shaded region shows the temperature range associated with Hallett-Mossop rime splintering, −3°C to −8°C.

Throughout this section, we analyze output from the ERA5-based LES from times overlapping the in situ observations used for model evaluation (second column in Table 2), which are always between 11 and 14 hr into the simulation. We analyze hours 10–12 from the 12 hr Obs-based simulations.

The second and fourth columns within each panel of Figure 7 show the differences between the simulated humidity and temperature profiles and the reference profiles. The Obs-based LES is not included here because its temperature and humidity are negligibly different from the reference profile, due to the strong nudging.

The stratocumulus cases, shown in the top row of Figure 7, feature well-mixed boundary layers topped by approximately adiabatic stratocumulus clouds. For RF12, the cloud layer occupies the Hallett-Mossop temperature range, whereas it is almost entirely above freezing for RF13. The GCMs and ERA5 have large temperature and humidity biases around temperature inversions topping the boundary layers due to vertical placement errors and inadequate vertical resolution. In the GCMs, this can lead to biases in entrainment,

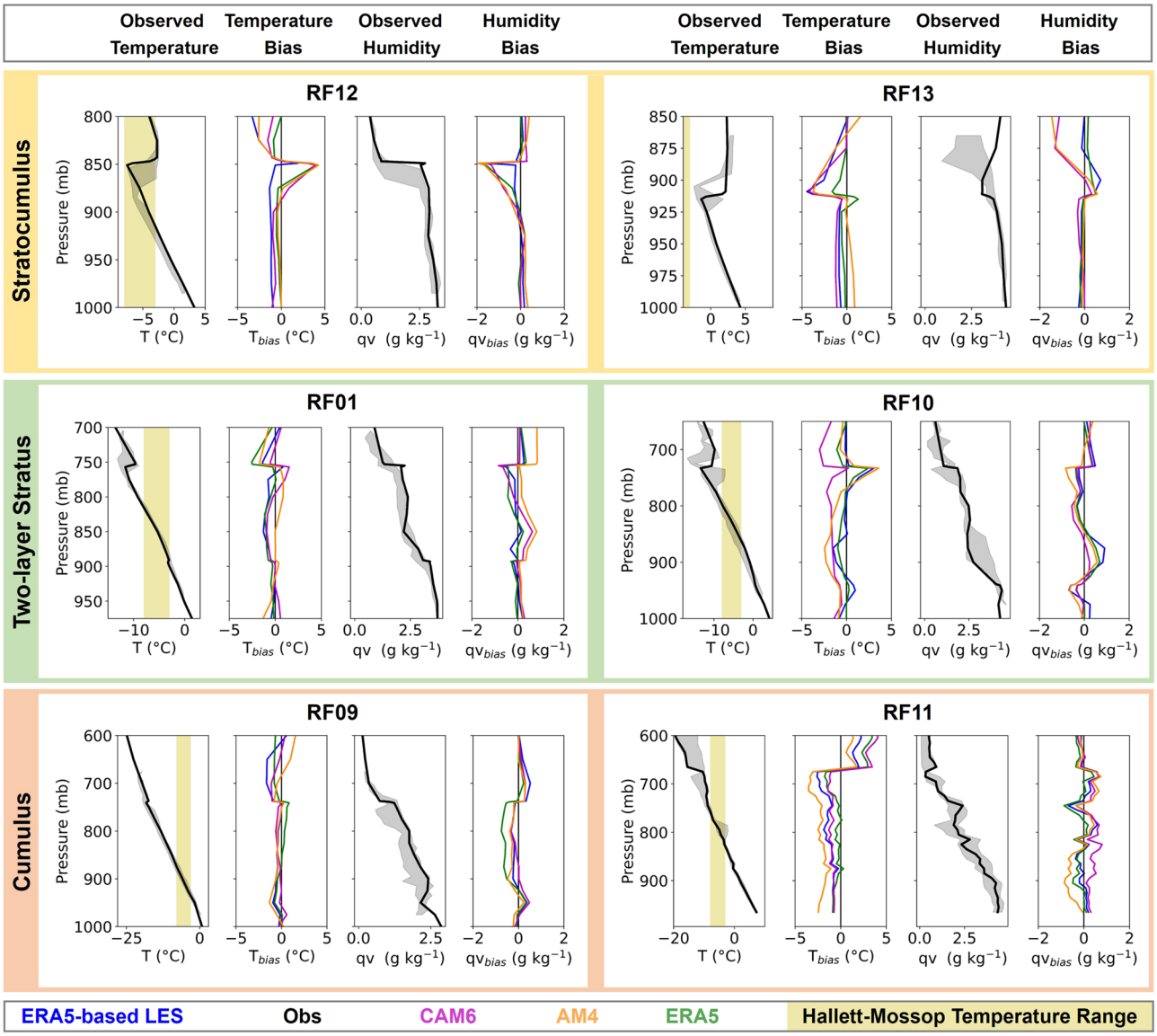

Figure 7. Within each panel, the first and third columns show the temperature and humidity from the in situ observations, respectively. The solid black line is the reference profile, to which the Obs-based LES is nudged, and the shaded gray area is the 10th to 90th percentile from the observations. The yellow shaded area marks the Hallett-Mossop temperature range. The second and fourth columns within each panel show profiles of median simulated temperature and humidity with the reference profiles subtracted. CAM6, AM4, and the ERA5-based LES are all nudged to ERA5 reanalysis.

an important control on boundary layer humidity and temperature, and can alter the vertical extent of the clouds. CAM6 develops cold biases throughout the boundary layer in both cases and AM4 develops a warm bias in RF12, possibly due to the diffuse inversion. 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 Appendix D.

The two-layer stratus cases, shown in the middle row of Figure 7, both have strong temperature inversions capping the upper cloud layers. They also both feature decoupled humidity profiles with sharp decreases above each cloud layer. The GCMs and ERA5 again show biases around the topmost inversion, due to their low vertical resolution. In RF01, CAM6 and AM4 do not simulate either the observed temperature inversion at 900 mb or the decoupled humidity profile and are moister than the observations and ERA5 between the two temperature inversions, as a result. In RF10, CAM6 and AM4 develop cold biases throughout the boundary layer that are larger than the cold bias in ERA5.

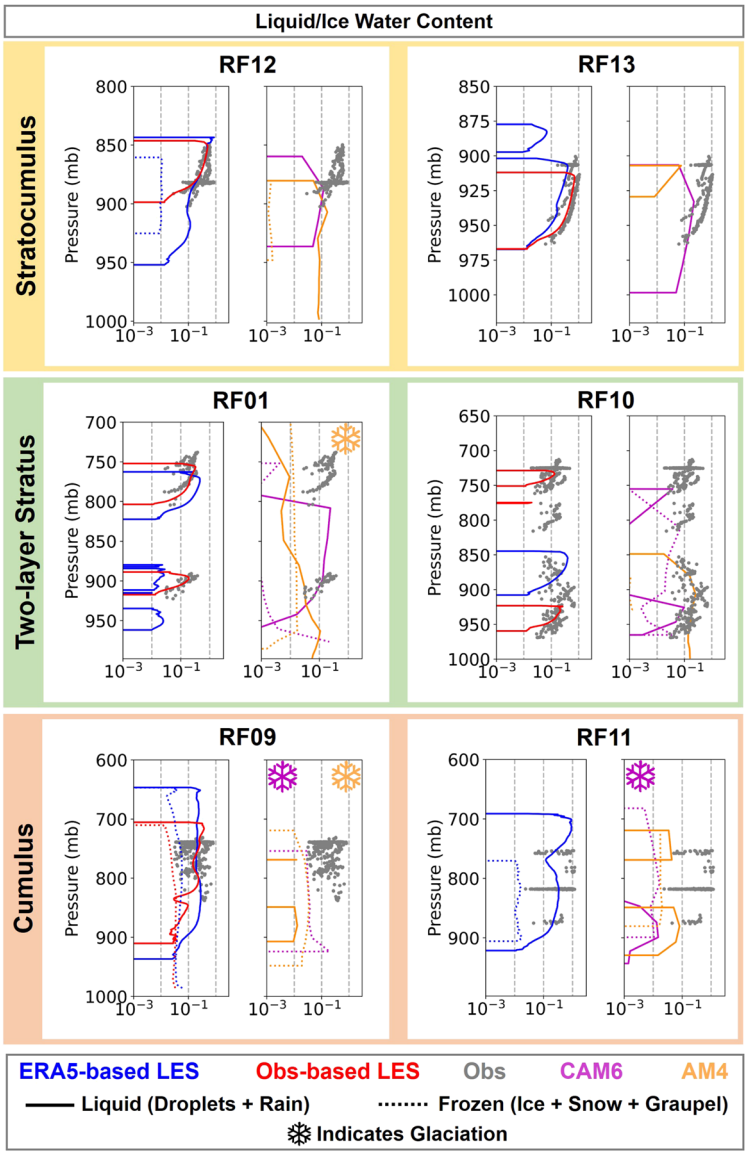

Figure 8. Binned in situ LWC from the CDP is shown with gray circles in both columns within each panel. Simulated profiles of median LWC and IWC are shown using solid lines and dotted lines, respectively. LES experiments are shown in the left column within each panel, and GCMs are shown on the right column within each panel. Snowflake icons indicate where there is glaciation in the GCMs, with the color indicating which GCM is being referenced.

Cumuliform cases are shown in the bottom row of Figure 7. RF09 features cumulus rising into stratocumulus, and RF11 features open cell cumulus. In RF11, the GCMs and the ERA5-based LES develop cold biases throughout the boundary layer, likely related to the smaller cold biases in ERA5 at the top of the boundary layer. ERA5 is warmer and drier than the LES or the GCMs in RF09, but all simulations are within with the range of the observations.

5.3. Cloud Macrophysics

SOCRATES clouds are dominated by supercooled water (Z2020). In this section, we compare profiles of in-cloud LWC between in situ observations, the LES experiments, and the GCMs. Within each panel of Figure 8, the left column shows the binned observations (gray dots) and the LES experiments, and the right column shows the binned observations and the GCMs.

We also show simulated profiles of ice water content (IWC), for which we do not have well-quantified measurements. Large frozen hydrometeors are qualitatively compared between observations and simulations using radar reflectivity in section 5.1 and using particle size distributions (PSDs) in section 5.6.

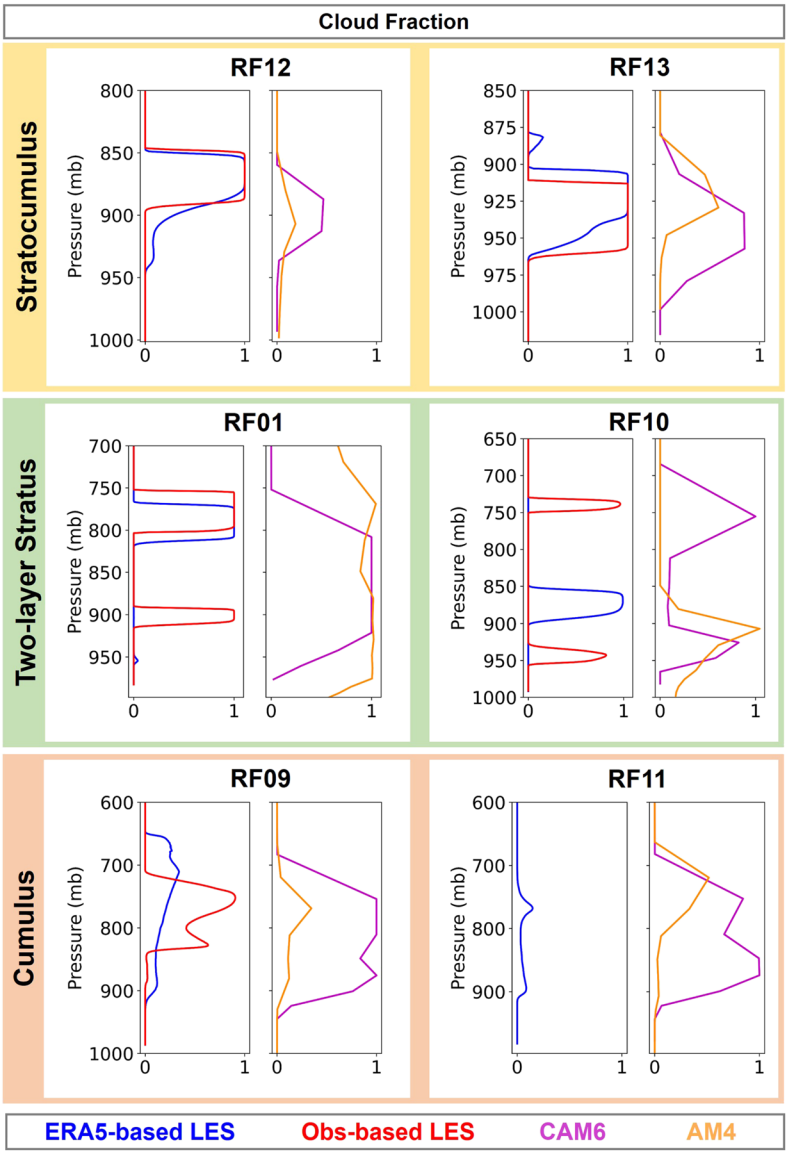

Figure 9. Median simulated cloud fraction is shown for the LES experiments in the left column within each panel and the GCMs in the right column within each panel.

All simulations produce supercooled water-dominated stratocumulus clouds for RF12 and RF13. The GCMs produce a cloud that is too low for RF12 and too thin in both cases. The ERA5-based LES produces a cloud that is too thin in RF13. For RF01, only the Obs-based LES is able to simulate two robust cloud layers. The ERA5-based LES skillfully simulates the upper cloud layer and has a small amount of liquid water in the location of the lower cloud but far less than the observed amount. The GCMs both simulate a single deep cloud instead of two cloud layers. AM4 simulates ice clouds between the two observed supercooled water-dominated cloud layers and above the top of the upper cloud. The low-resolution models struggle similarly with RF10, with the exception of CAM6, which simulates two robust supercooled water-dominated cloud layers, with thinner ice cloud in between.

CAM6 simulates an ice-dominated cloud in both cumuliform cases (bottom row of Figure 8). The cumulus parameterization turns on in these two cases and in RF10 and likely contributes to the excessive glaciation. AM4 simulates an ice-dominated cloud in RF09 but maintains supercooled water throughout the boundary layer in RF11.

Cloud fraction and cloud thickness together determine the radiative properties of boundary layer clouds so we also compare cloud fraction between the simulations in Figure 9, where the LES experiments are

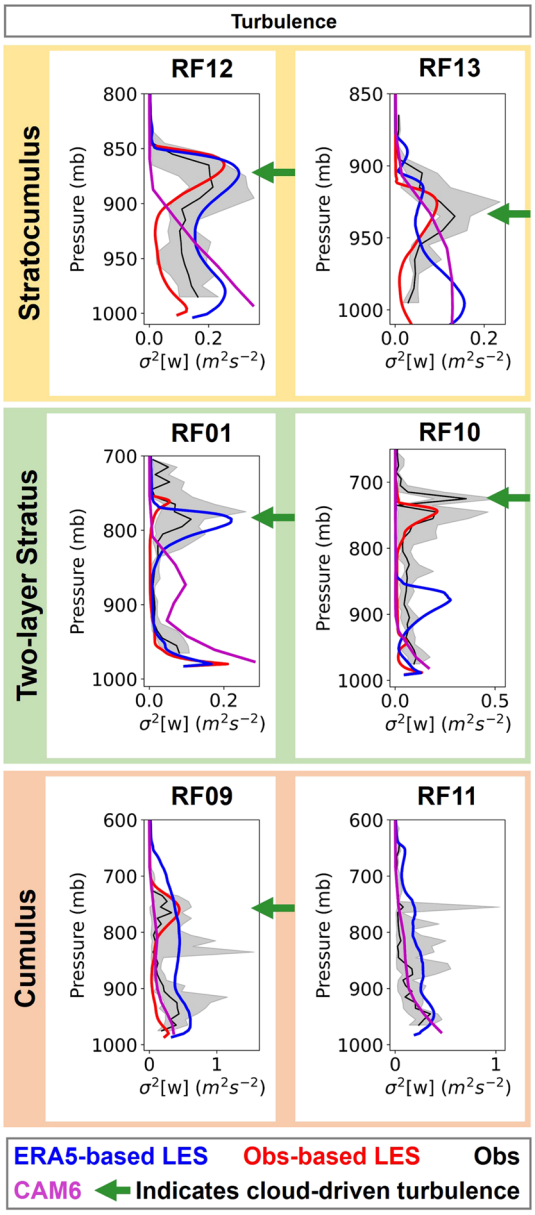

Figure 10. Vertical wind variance is compared between the in situ observations, the LES experiments, and CAM6 within each panel. The black line is the median of the observations, and the gray shaded area is the 10th to 90th percentile. All simulated profiles are medians. Green arrows indicate vertical layers with observed peaks in turbulence, driven by longwave cooling at the cloud top.

in the left column within each panel and the GCMs are in the right column. The LES cloud fraction is a domain average. The radar profiles in Figure 6 provide a qualitative comparison with observations. The stratocumulus clouds in RF09, RF12, and RF13 and the upper cloud layers in RF01 and RF10 are persistent throughout their respective modules and have a cloud fraction close to 1. The lower cloud in RF01 is slightly less persistent, and the lower cloud layer in RF10, as well as the cumuli in RF11, is intermittent. The LES consistently simulates a cloud fraction of 1 in the stratiform cloud layers (top and middle rows of Figure 9). The ERA5-based LES underestimates stratocumulus cloud cover in RF09, possibly related to the boundary layer deepening excessively. The relatively small 12.8 km square model domain does not permit the development of the largest cumuli, which can have cross sections of up to 10 km, possibly biasing cumulus cloud cover low in RF09 and RF11.

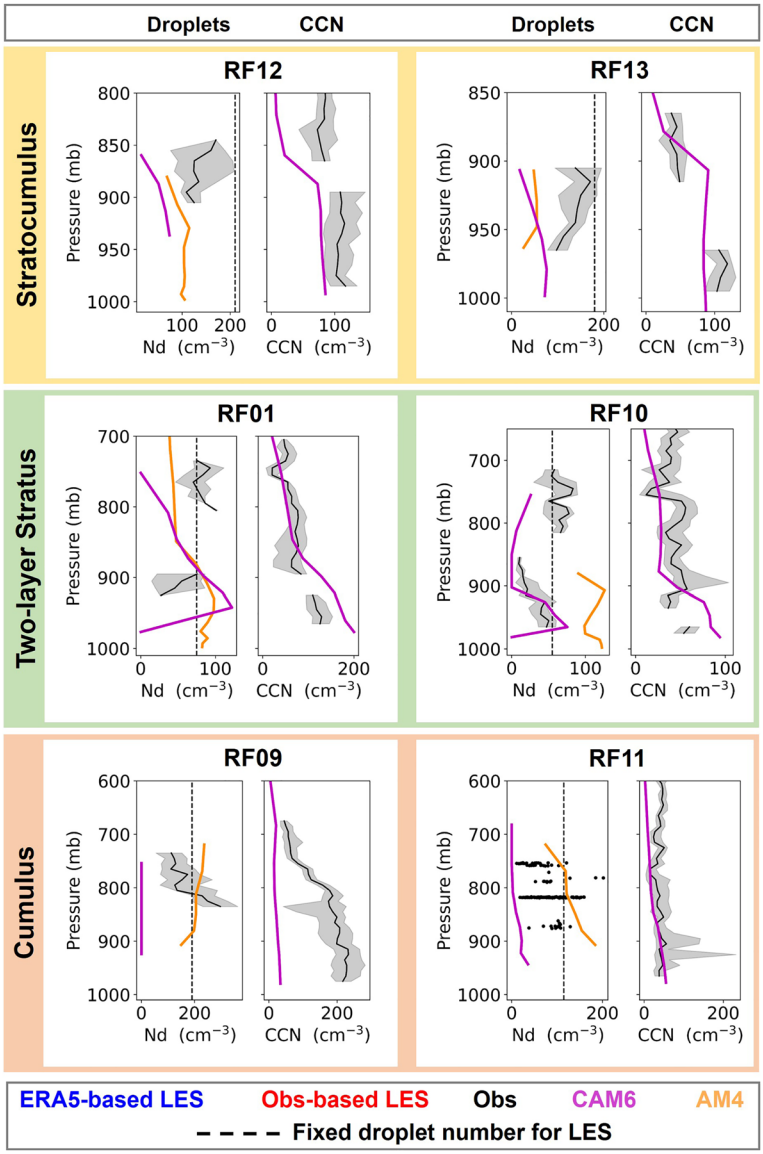

Figure 11. Droplet number concentrations are compared between CAM6, AM4, and observations from the CDP in the left column within each panel. The vertical dashed line is the fixed droplet concentration used for the LES experiments. CCN concentrations from CAM6 (at 0.2% supersaturation) are compared with the large particles measured by the UHSAS in the right column within each panel. In all plots, the black line is the median of the observations and the gray shaded area is the 10th to 90th percentile. All simulated profiles are medians.

The GCMs, and AM4 in particular, have too little cloud cover in both stratocumulus cases (RF12 and RF13), whereas CAM6 has too much cloud cover in both cumulus cases (RF09 and RF11).

5.4. Turbulence

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, latent heating and evaporation, and longwave cooling at the cloud top. SOCRATES boundary layers are usually characterized by decoupled turbulence profiles with distinct peaks near the sea surface and within each cloud layer. Multilayer stratus clouds and cumulus are associated with stronger decoupling.

We compare turbulence between the observations, the LES experiments, and CAM6, using profiles of vertical wind variance, in Figure 10. For the LES, we estimate the total vertical wind variance by adding the

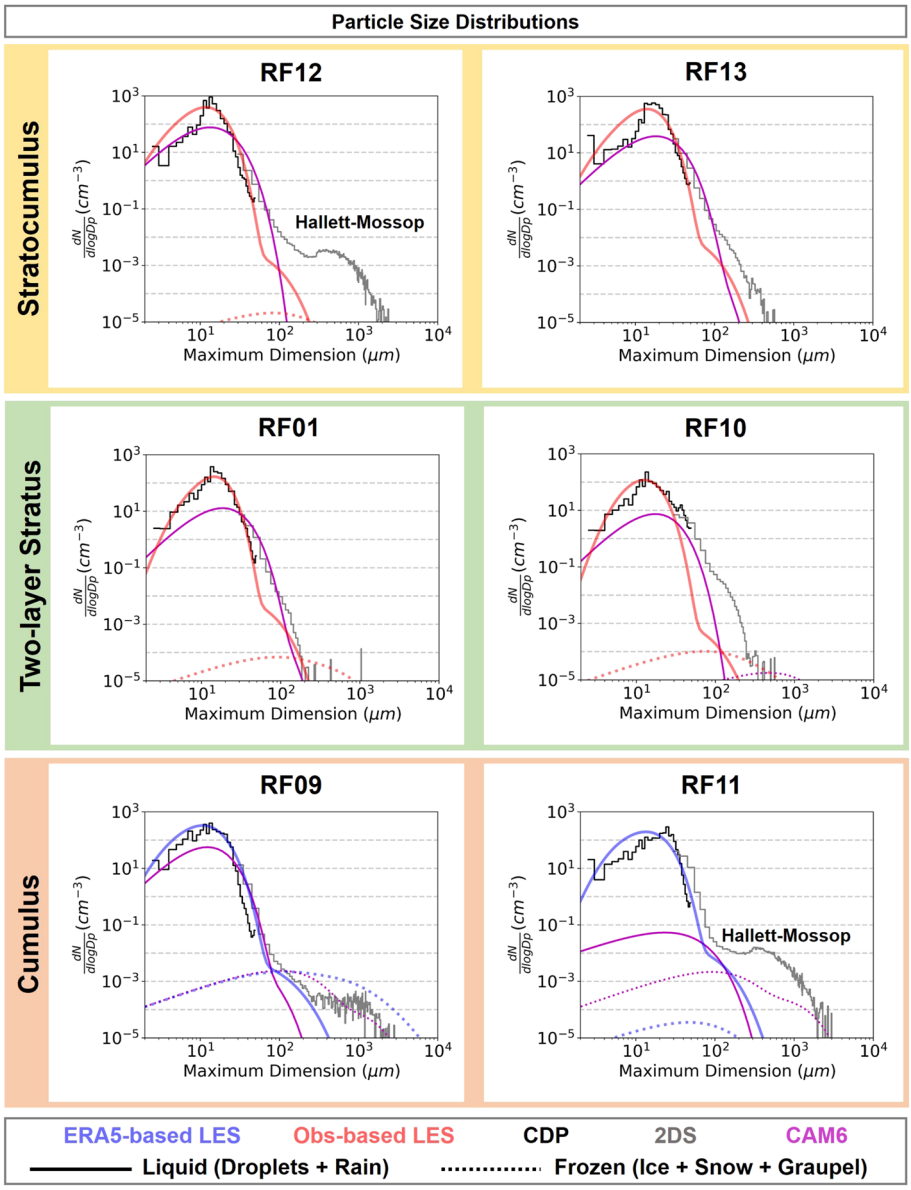

Figure 12. Particle size distributions from the Cloud Droplet Probe (CDP) (thin black line), Two-Dimensional Stereo Probe (2DS) (thin gray line), the LES (blue or red), and CAM6 (purple) are shown for each case. Red lines are the Obs-based LES, and blue lines are the ERA5-based LES. For the simulations, solid lines are liquid hydrometeor PSDs and dotted lines are frozen hydrometeor PSDs.

resolved vertical wind variance and two thirds of the subgrid-scale turbulent kinetic energy (i.e., equipartitioning of subgrid turbulent kinetic energy between coordinate directions) (Morrison & Pinto, 2005). In general, the resolved vertical wind variance dominates the subgrid contribution. We do not have vertical wind variance from AM4.

In all four stratiform cases (top and middle rows of Figure 10) and in the stratocumulus cloud deck in RF09 (lower left panel in Figure 10), turbulent peaks that are not adjacent to the sea surface are vertically collocated with peaks in LWC, in the observations, and in both LES experiments. Simulated profiles of cloud radiative heating from the LES (not shown) reveal strong longwave cooling at the top of the uppermost stratiform cloud layers. The lower observed cloud layers in RF01 and RF10 (middle row of Figure 10) are not associated with peaks in turbulence because they are masked by the upper cloud layers and experience less longwave cooling (Sedlar & Shupe, 2014).

CAM6 lacks in-cloud enhancement of turbulence in four out of the five cases where it is observed (RF12, RF13, RF10 and RF09). CAM6 simulates turbulence within the cloud layer in RF01, but it is not concentrated at the top of the cloud at 810 mb (Figure 8) and is unlikely to be radiatively driven.

RF01 and RF10 feature stable boundary layers and have localized peaks in turbulence near the sea surface, due to wind shear. Surface heat fluxes are not a dominant source of turbulence in any of the stratiform cases.

On the other hand, in the cumuliform cases (bottom row of Figure 10), thermal instability at the sea surface and cumulus cloud formation create a robust peak in the turbulence within the lower part of the boundary layer. The ERA5-based LES and CAM6 capture the turbulence associated with rising cumuli, but the Obs-based LES misses this contribution in RF09 due to not resolving the rising cumuli.

5.5. CCN and Droplet Number Concentration

We use in situ observations to evaluate droplet concentrations and Cloud Condensation Nuclei (CCN) concentrations in CAM6, in the left and right columns within each panel in Figure 11, respectively. We calculate the observed droplet concentrations from the Cloud Droplet Probe (CDP), and we use the observed concentration of particles greater than 0.1 μm from the Ultra-High Sensitivity Aerosol Spectrometer (UHSAS) as a proxy for CCN. There were also two CCN counters on the G-V, but one was scanning through a wide range of supersaturations at all times, and the other had frequent problems leading to missing data. When UHSAS and CCN measurements at a supersaturation of 0.2% could be obtained simultaneously, they typically agreed well (not shown). For CAM6, CCN at 0.2% supersaturation was diagnosed.

We also include droplet concentrations from AM4, but we do not have CCN concentrations from that model. The LES uses fixed droplet concentrations, which are specified in Table 2, and shown in the left column of each panel in Figure 11, using dashed vertical lines.

CAM6 is deficient in droplets in all cases except within the lower cloud layers in RF01 and RF10, where there is substantial shear-driven turbulence. AM4 has too few droplets in the two stratocumulus cases (top row of Figure 11) but otherwise does not exhibit a low droplet concentration bias. We consider three factors that may contribute to low droplet concentrations in CAM6: deficient CCN, excessive glaciation, and deficient cloud-driven turbulence. Turbulent vertical motions can affect droplet concentrations in two ways: directly, by modulating the activation of CCN (Abdul-Razzak & Ghan, 2000), and indirectly, by vertically mixing CCN within the boundary layer and entraining CCN from above the boundary layer.

The low droplet concentration bias is particularly stark in RF09 and RF11, even though the turbulence in CAM6 agrees well with observations throughout the cloud layers. RF09 is deficient in CCN but RF11 is not. Both cases feature excessive glaciation, so we posit that to be the primary cause of their low droplet concentrations.

The other four stratiform cases do not feature excessive glaciation and are not deficient in CCN. However, they all have inadequate stratiform cloud-driven turbulence; we posit that this is the main reason for CAM6 underestimating cloud droplet concentrations in these cases.

5.6. PSDs

Cloud hydrometeor properties influence cloud lifetime, cloud radiative effects, and precipitation. To evaluate simulated hydrometeors, we compare PSDs between in situ observations and simulations in Figure 12. AM4 hydrometeors are one moment so AM4 is left out of this comparison. We compare CAM6 with the Obs-based LES for the four stratiform cases and the ERA5-based LES for the two cumuliform cases.

We have separate PSDs for each hydrometeor from the simulations, but it is 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 during SOCRATES at temperatures of −5°C 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 small ice crystals.

RF11 and RF12 (upper left and bottom right plots in Figure 12) feature clouds within the Hallett-Mossop temperature range and have large particle modes ($200 < D < 2000$ μm) which we assume represent predominantly frozen particles. Both CAM6 and the Obs-based LES fail to reproduce the observed large particle

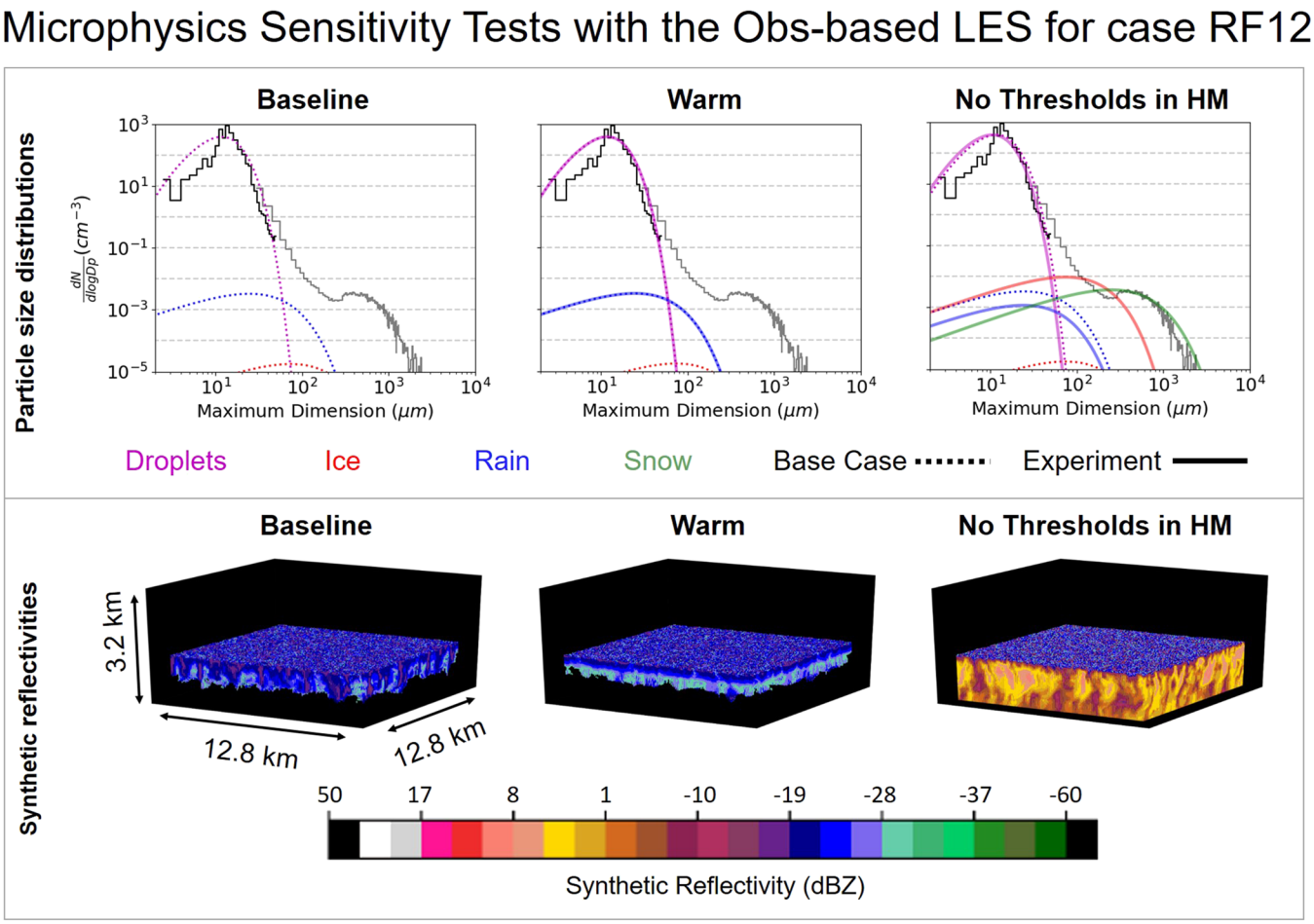

Figure 13. PSDs (top row) and synthetic reflectivities (bottom row) are shown for the baseline Obs-based simulation of RF12 (left column) and the two microphysics sensitivity tests (right columns). The dashed lines in all plots in the top row show the baseline PSDs, and the solid lines show the PSDs from the sensitivity tests.

modes for RF12 and CAM6 lacks frozen particles entirely. The ERA5-based LES also fails to reproduce the large particle mode for RF11, whereas CAM6 does simulate one but it contains fewer particles than observed and is achieved through excessive glaciation. Thus, neither model is able to generate the observed high concentration of large frozen particles within a supercooled water-dominated cloud occupying the Hallett-Mossop temperature range. We perform sensitivity tests to Hallett-Mossop rime splintering in the LES in Section 6.

The Obs-based LES is deficient in mid-sized particles ($50 < D < 200$ μm), which could be drizzle or small ice, in all cases except RF09. The low cloud droplet concentration bias in CAM6 leads to larger droplets and more efficient autoconversion of droplets to rain. As a result, CAM6 simulates more of these drizzle droplets than the LES in all four stratiform cases (top and middle rows of Figure 12). This suggests that autoconversion may be too inefficient for small droplets (D ~ 10 μm), which dominate SOCRATES clouds.

6. LES Sensitivity to Primary and Secondary Ice Production

In this section, we present two microphysics 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.

The left column of Figure 13 shows the PSD and synthetic radar reflectivity for this baseline simulation. In contrast to the observations (Figures 6 and 12), the PSD shows almost no large particles, and the reflectivity is correspondingly weak.

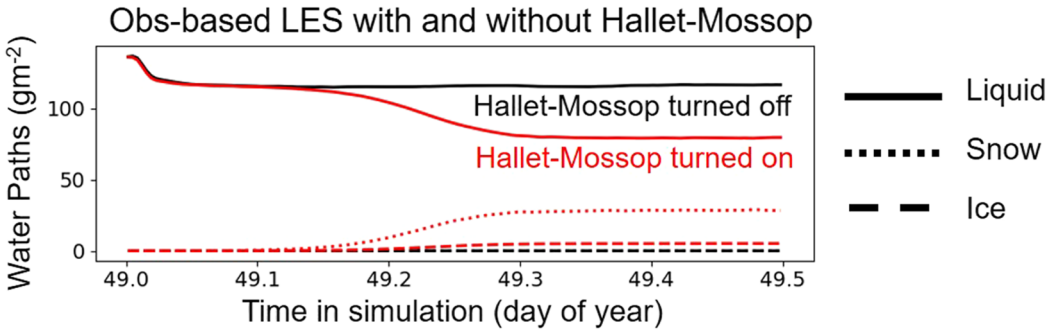

Figure 14. Time series of liquid, snow, and ice water paths are shown for the baseline Obs-based simulation (black) and the simulation with the thresholds in the Hallett-Mossop parameterization removed (red).

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 (middle column of Figure 13).

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 process enables the LES to skillfully reproduce the observed large particle mode. It also initiates precipitation in the LES and drastically increases the simulated reflectivities (right column of Figure 13), which now resemble the highest reflectivities from the observations (second to bottom row of Figure 6). Ice and snow concentrations increase by 2–3 orders of magnitude, while there is a modest reduction of cloud and drizzle droplets and a 30% decrease in the liquid water path (Figure 14), part of which is due to the Hallett-Mossop process reducing the number of cloudy grid cells and is not captured by the in-cloud PSDs.

The MG2008 microphysics scheme in CAM6 does not have any mass thresholds within the Hallett-Mossop rime splintering parameterization. Still, CAM6 is not able to reproduce the large particle mode for RF12 because it does not form any ice through primary nucleation (Figure 12), which is a necessary precursor to Hallett-Mossop rime splintering.

Our results differ from Young et al. (2019), who achieved the best agreement between simulated PSDs and 2DS observations when they removed only the liquid threshold in the M2005 rime splintering parameterization and increased the multiplication factor (number of splinters per milligram of rime accreted) by an order of magnitude.

Laboratory studies of the Hallett-Mossop process have converged on a set of necessary conditions for rime splintering: large frozen particles to serve as rimers, small droplets (<13 μm), which retain their shape upon freezing on the rimers, and large droplets (>25 μm), which splinter when they come into contact with the small frozen droplets on the surface of the rimers (Field et al., 2016; Hallett & Mossop, 1974). The mass thresholds in M2005 have no clear relationship to these conditions.

All SOCRATES LES experiments feature small and large droplets throughout the supercooled water-dominated clouds (Figure 12), and all cases but RF13 have small concentrations of frozen particles, thus fulfilling the requirements for Hallett-Mossop rime splintering, given the correct temperature range. However, SOCRATES LES baseline experiments rarely have LWC > 0.5 g kg^{-1} (Figure 8) and never have rain mass > 0.1 g kg^{-1} (not shown) or frozen hydrometeor mass > 0.1 g kg^{-1} (Figure 8). Thus, we recommend removing all of the mass thresholds in the M2005 Hallett-Mossop rime splintering parameterization in simulations of Southern Ocean boundary layer clouds.

7. Conclusions

The SOCRATES campaign sampled stratiform and cumuliform 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 supercooled water-dominated Southern Ocean boundary layer clouds.

SAM LES maintains supercooled water in all cases and skillfully reproduces the observed turbulence in stratiform cloud layers. We developed two complementary forcing strategies to create the flexibility to simulate diverse Southern Ocean cloud morphologies. The Obs-based LES works well for simulating multiple stratiform cloud layers because it is forced to maintain thin cloud layers and boundary layer decoupling associated with small-scale features in the observed temperature and humidity profiles that are not captured by the ERA5 reanalysis. The ERA5-based LES works well for simulating cumuliform clouds for which strong nudging to a single thermodynamic profile cannot capture the horizontal inhomogeneity of the cloud field.

The LES simulates mixed-phase cloud microphysics fairly well but with two notable biases. First, the LES has too few midsize particles ($50 < D < 200$ μm diameter) in all but one case and second, it produces too little snow and graupel ($200 < D < 2000$ μm diameter) in cases which feature clouds within the temperature range that Hallet-Mossop is active, biasing the precipitation low. Sensitivity tests using case RF12 showed that the second bias can be rectified by removing thresholds in the M2005 microphysics that inactivate Hallett-Mossop rime splintering.

The two nudged GCMs, CAM6 and AM4, also correctly predict that Southern Ocean supercooled stratiform boundary layer clouds are liquid dominated in all considered SOCRATES cases. The vertical placement of cloud layers in both GCMs is less realistic than the LES. They achieve the most realistic cloud morphology for the two stratocumulus cases but have a lower cloud fraction than observed. They do less well on the two two-layer stratiform cases because they cannot maintain the sharp inversions and weak boundary layer decoupling observed in these cases.

Increased vertical resolution might help, but even the ERA5 reanalysis, to which the GCM temperature profiles are nudged, does not capture key details of the observed boundary layer structure in these cases.

In CAM6, boundary layer cloud droplet concentrations are systemically too low. One reason is that deep convection is easily triggered in shallow cumulus-forming environments, leading to excessive glaciation and low droplet concentrations. Second, inadequate turbulent vertical wind variance simulated in cloud layers by the CLUBB parameterization limits aerosol activation into cloud droplets.

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 evaluate simulated clouds and cloud processes against the real world. The SOCRATES data set 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. We encourage other modeling teams to use the SOCRATES data set and the LES cases presented here to methodically evaluate and improve simulations of Southern Ocean boundary layer clouds from a process-level perspective.

Appendix A: Correction of the Observed 20 s Running Vertical Wind Variance

To correct $\sigma^2_{20}[w]$ for the portion of the true vertical wind variance that occurs on scales larger than 20 s, we constructed a power spectrum of w for each below-cloud and in-cloud leg from SOCRATES, and we computed the fraction of the vertical wind variance associated with periods greater than 20 s (f_{20}), which we show in Figure A1. Since this fraction increases with altitude (z, in meters), we made the following altitude-dependent correction to $\sigma^2_{20}[w]$ throughout our study to obtain an estimate of the full vertical wind variance ($\widehat{\sigma^2[w]}$) which can be directly compared with model-derived vertical wind variance:

$$\widehat{\sigma^2[w]} = \frac{\sigma^2_{20}[w]}{1 - f_{20}}, \tag{A1}$$

$$f_{20}(z) = 0.167 + 1.267 \times 10^{-4} \min(z, 2000). \tag{A2}$$

Equation A2 is equal to the red line in Figure A1.

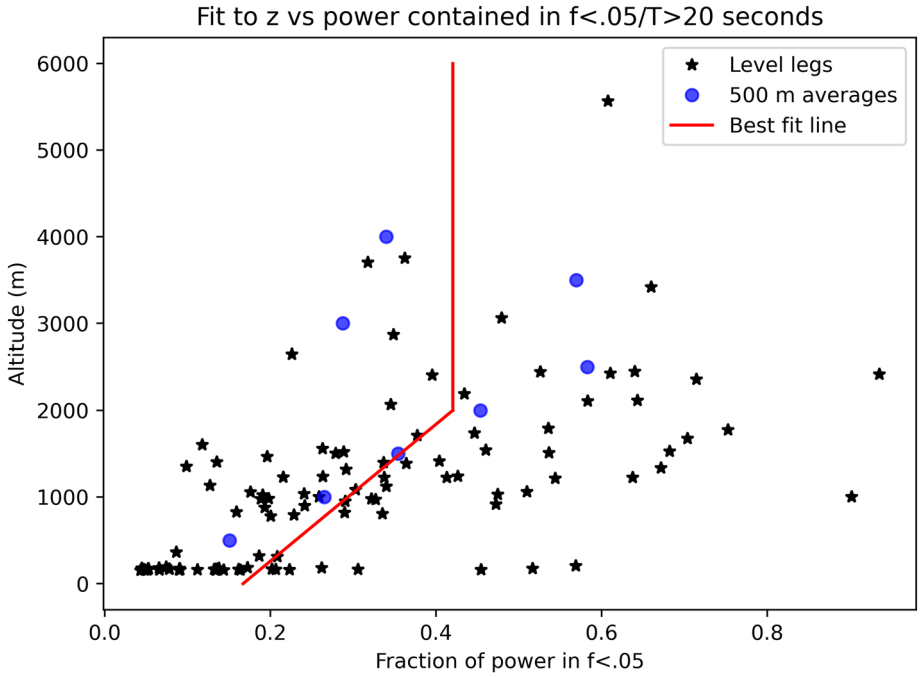

Figure A1. The fraction of power contained in periods greater than 20 s is shown for each horizontal flight leg from SOCRATES, as a function of the mean altitude of the flight leg (black stars). Blue dots show average values from 500 m thick bins, and the red line is the fit to the black stars, which is used to correct the vertical wind variance.

Appendix B: Preparation of the Obs-Based Model Input

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 vertical-cavity surface-emitting laser (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. 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 than to a profile that has false deviations from an adiabatic profile due to covering a large horizontal area (~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.

Appendix C: Estimation of Vertical Wind From ERA5-Reported Pressure Velocity

We estimate the large-scale vertical wind from the ERA5-reported pressure velocity based on an approximate formula, valid near the surface:

$$\begin{aligned}\omega(x,y,p,t) &= \frac{D_p p}{Dt} = \frac{D_p p_s}{Dt} + \frac{D_p}{Dt}(p-p_s)\\ &= \frac{D_s p_s}{Dt} + (u-u_s)\frac{\partial p_s}{\partial x} + (v-v_s)\frac{\partial p_s}{\partial y} + \frac{D_p}{Dt}(p-p_s)\\ &\approx \omega_s + w\frac{\partial p}{\partial z}.\end{aligned} \tag{C1}$$

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. D_p/Dt and D_s/Dt are different because there are different horizontal winds at p_s and p. 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:

$$\omega \approx f(p)\omega_s + w\partial p/\partial z, \tag{C2}$$

where $f(p)$ is a sigmoidal curve that is equal to 1 at the surface and decays to 0 at the tropopause (250 mb). Using the hydrostatic approximation to calculate the vertical pressure gradient, we calculate vertical velocity from ERA5 as follows:

$$w \approx -\frac{\omega - \omega_s f(p)}{\rho g}. \tag{C3}$$

Appendix D: Explanation of Free Tropospheric Cold Biases in ERA5-Based LES

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 (Figure 7). Southern Ocean stratocumulus cases are often associated with strong, rapidly evolving temperature inversions at the top of the boundary layer. Figure D1a 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 preexisting inversion at 800 mb and ends up with two inversions, which develop a large cold bias between them (Figure D1b). This midtropospheric 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 modest nudging with a 12 hr timescale 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.

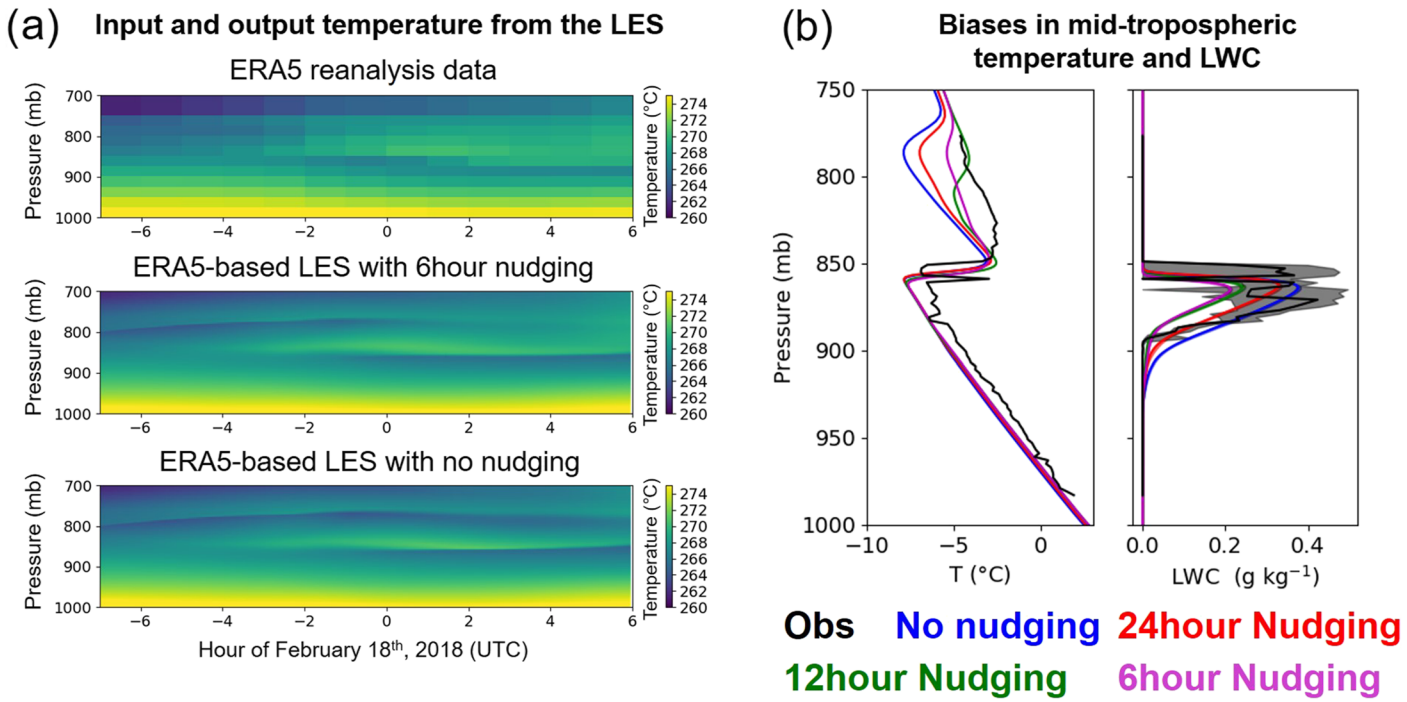

Figure D1. The left column (a) shows temperature from pressure-level ERA5 data (top), temperature from the ERA5-based LES with 6 hr nudging (middle), and temperature from the ERA5-based LES with no nudging (bottom). The right column (b) shows profiles of temperature (left) and LWC (right) for different nudging timescales.

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.

Data Availability Statement

The data used in this study rely on a collaborative effort by a broad, dedicated SOCRATES team led by Greg McFarquhar. All aircraft data are publicly available online (https://data.eol.ucar.edu/master_lists/generated/socrates/). SAM LES model input and output for the cases described here are publicly available online (https://atmos.uw.edu/~ratlas/SOCRATES-LES-cases.html).

Acknowledgments
The authors gratefully acknowledge funding from U. S. National Science Foundation (NSF) grants AGS-1660604 (UW and NCAR) and AGS-1660609 (UW), AGS-1660538 (GFDL), and an NSF Graduate Student Research Fellowship to R. Atlas (DGE-1762114). We thank all the SOCRATES scientists and NCAR 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.

References

Abdul-Razzak, H., & Ghan, S. J. (2000). A parameterization of aerosol activation: 2. Multiple aerosol types. *Journal of Geophysical Research*, *105*(D5), 6837–6844. https://doi.org/10.1029/1999JD901161

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

Berner, A. H., Bretherton, C. S., & Wood, R. (2015). Large eddy simulation of ship tracks in the collapsed marine boundary layer: A case study from the Monterey area ship track experiment. *Atmospheric Chemistry and Physics*, *15*(10), 5851–5871. https://doi.org/10.5194/acp-15-5851-2015

Bodas-Salcedo, A., Hill, P. G., Furtado, K., Williams, K. D., Field, P. R., Manners, J. C., et al. (2016). Large contribution of supercooled liquid clouds to the solar radiation budget of the southern ocean. *Journal of Climate*, *29*(11), 4213–4228. https://doi.org/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*, 1735–1758. https://doi.org/10.1029/2019ms001688

Bodas-Salcedo, A., Williams, K. D., Ringer, M. A., Beau, I., Cole, J. N. S., Dufresne, J. L., et al. (2014). Origins of the solar radiation biases over the southern ocean in CFMIP2 models. *Journal of Climate*, *27*(1), 41–56. https://doi.org/10.1175/jcli-d-13-00169.1

Bretherton, C., McCaa, J., & Grenier, H. A. (2004). A new parameterization for shallow cumulus convection and its application to marine subtropical Cloud-Topped boundary layers. Part I: Description and 1D results. *Monthly Weather Review*, *132*(4), 864–882. https://doi.org/10.1175/1520-0493(2004)132(0864:ANPFSC)2.0.CO;2

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., et al. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature*, *503*(7474), 67–71. https://doi.org/10.1038/nature12674

Cooper, W. A., Friesen, R. B., Hayman, M., Jensen, J. B., Lenschow, D. H., Romashkin, P. A., et al. (2016). Characterization of uncertainty in measurements of wind from the NSF/NCAR gulfstream v research aircraft (*NCAR/TN-528+STR*): NCAR Technical Notes.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The community Earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001916. https://doi.org/10.1029/2019ms001916

DeMott, P. J., Hill, T. C. J., McCluskey, C. S., Prather, K. A., Collins, D. B., Sullivan, R. C., et al. (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. https://doi.org/10.1073/pnas.1514034112

Field, P. R., Lawson, R. P., Brown, P. R. A., Lloyd, G., Westbrook, C., Moisseev, D., et al. (2016). Chapter 7. secondary ice production—Current state of the science and recommendations for the future. *Meteorological Monographs*, *58*, D–16–0014.1. https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0014.1

Gettelman, A., Bardeen, C. G., McCluskey, C. S., Järvinen, E., Stith, J., Bretherton, C., et al. (2020). Simulating observations of southern ocean clouds and implications for climate. *Journal of Geophysical Research: Atmospheres*, *125*, e2020JD032619. https://doi.org/10.1029/2020JD032619

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al. (2019). High climate sensitivity in the community earth system model version 2 (CESM2). *Geophysical Research Letters*, *46*, 8329–8337. https://doi.org/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*, 4540–4547. https://doi.org/10.1002/2015GL063672

Hallett, J., & Mossop, S. C. (1974). Production of secondary ice particles during riming process. *Nature*, *249*(5452), 26–28. https://doi.org/10.1038/249026a0

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., et al. (2018). Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nature Communications*, *9*, 3625. https://doi.org/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. https://doi.org/10.1175/1520-0469(2003)060<0607:crmota>2.0.co;2

LeMone, M. A., & Zipser, E. J. (1980). Cumulonimbus vertical velocity events in GATE. Part I: Diameter, intensity and mass flux. *Journal of the Atmospheric Sciences*, *37*(11), 2444–2457. https://doi.org/10.1175/1520-0469(1980)037<2444:cvveig>2.0.co;2

Liu, X., Ma, P. L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., et al. (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. https://doi.org/10.5194/gmd-9-505-2016

McCluskey, C. S., Hill, T. C. J., Humphries, R. S., Rauker, A. M., Moreau, S., Strutton, P. G., et al. (2018). Observations of ice nucleating particles over Southern Ocean waters. *Geophysical Research Letters*, *45*, 11,989–11,997. https://doi.org/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*, 650–668. https://doi.org/10.1002/2015MS000589

McFarquhar, G., Bretherton, C., Marchand, R., Protat, A., Demott, P., Alexander, S., et al. (2020). Observations of clouds, aerosols, precipitation, and surface radiation over the southern ocean: An overview of CAPRICORN, MARCUS, MICRE and SOCRATES. *Bulletin of the American Meteorological Studies.*

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. https://doi.org/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. https://doi.org/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*, *102*(D14), 16,663–16,682. https://doi.org/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. https://doi.org/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. https://doi.org/10.1175/2008JCLI2105.1

Morrison, H., & Pinto, J. O. (2005). Mesoscale modeling of springtime arctic Mixed-Phase stratiform clouds using a new two-moment bulk microphysics scheme. *Journal of the Atmospheric Sciences*, *62*(10), 3683–3704. https://doi.org/10.1175/jas3564.1

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. https://doi.org/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. https://doi.org/10.1038/nclimate2513

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

Schmidt, K. M., Swart, S., Reason, C., & Nicholson, S.-A. (2017). Evaluation of satellite and reanalysis wind products with in situ wave glider wind observations in the southern ocean. *Journal of Atmospheric and Oceanic Technology*, *34*(12), 2551–2568. https://doi.org/10.1175/jtech-d-17-0079.1

Sedlar, J., & Shupe, M. D. (2014). Characteristic nature of vertical motions observed in Arctic mixed-phase stratocumulus. *Atmospheric Chemistry and Physics*, *14*(7), 3461–3478. https://doi.org/10.5194/acp-14-3461-2014

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. https://doi.org/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. https://doi.org/10.1126/science.aad5300

Tomassini, L., Field, P. R., Honnert, R., Malardel, S., McTaggart-Cowan, R., Saitou, K., et al. (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*, 39–64. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1175/mwr-d-10-05044.1

Young, G., Lachlan-Cope, T., O'Shea, S. J., Dearden, C., Listowski, C., Bower, K. N., et al. (2019). Radiative effects of secondary ice enhancement in coastal Antarctic clouds. *Geophysical Research Letters*, *46*, 2312–2321. https://doi.org/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*, *114*, D04106. https://doi.org/10.1029/2008JD010874

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*, e85782. https://doi.org/10.1029/2019GL085782

Zhao, M., Golaz, J. C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al. (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*, 735–769. https://doi.org/10.1002/2017MS001209

Zhou, X. L., Atlas, R., McCoy, I. L., Bretherton, C. S., Bardeen, C., Gettelman, A., et al. (2020). Evaluation of cloud and precipitation simulations in CAM6 and AM4 using observations over the Southern Ocean, 65. https://doi.org/10.1029/2020EA001241

Zhou, X. L., & Bretherton, C. S. (2019). Simulation of mesoscale cellular convection in marine stratocumulus: 2. Nondrizzling conditions. *Journal of Advances in Modeling Earth Systems*, *11*, 3–18. https://doi.org/10.1029/2018MS001448