

# Evaluation of cloud and precipitation simulations in CAM6 and AM4 using observations over the Southern Ocean

Zhou Xiaoli<sup>1</sup>, Atlas Rachel<sup>1</sup>, McCoy Isabel L.<sup>1</sup>, Bretherton Christopher S.<sup>1</sup>, Bardeen Charles<sup>2</sup>, Gettelman Andrew<sup>2</sup>, Lin Pu<sup>3</sup>, and Ming Yi<sup>4</sup>

<sup>1</sup>University of Washington

<sup>2</sup>National Center for Atmospheric Research (UCAR)

<sup>3</sup>Princeton University

<sup>4</sup>Geophysical Fluid Dynamics Laboratory

November 16, 2022

## Abstract

This study uses cloud and radiative properties collected from in-situ and remote sensing instruments during two coordinated campaigns over the Southern Ocean between Tasmania and Antarctica in January-February 2018 to evaluate the simulations of clouds and precipitation in nudged-meteorology simulations with the CAM6 and AM4 global climate models sampled at the times and locations of the observations. Fifteen SOCRATES research flights sampled cloud water content, cloud droplet number concentration, and particle size distributions in mixed-phase boundary-layer clouds at temperatures down to -25 C. The six-week CAPRICORN2 research cruise encountered all cloud regimes across the region. Data from vertically-pointing 94 GHz radars deployed was compared with radar-simulator output from both models. Satellite data was compared with simulated top-of-atmosphere (TOA) radiative fluxes.

Both models simulate observed cloud properties fairly well within the variability of observations. Cloud base and top in both models are generally biased low. CAM6 overestimates cloud occurrence and optical thickness while cloud droplet number concentrations are biased low, leading to excessive TOA reflected shortwave radiation. In general, low clouds in CAM6 precipitate at the same frequency but are more homogeneous compared to observations. Deep clouds are better simulated but produce snow too frequently.

AM4 underestimates cloud occurrence but overestimates cloud optical thickness even more than CAM6, causing excessive outgoing longwave radiation fluxes but comparable reflected shortwave radiation. AM4 cloud droplet number concentrations match observations better than CAM6. Precipitating low and deep clouds in AM4 have too little snow. Further investigation of these microphysical biases is needed for both models.



25 **Key points:**

- 26 1. CAM6 and AM4 simulate observed cloud properties and compositions fairly well  
27 within the variability of observations.
- 28 2. CAM6 clouds are “too frequent, too bright”; AM4 clouds are “too few, too  
29 bright”.
- 30 3. Cloud droplet number concentration in CAM6 is typically too low; AM4 clouds  
31 include too much small ice and too little snow.

32

33

34 **Abstract**

35

36 This study uses cloud and radiative properties collected from in-situ and remote sensing  
37 instruments during two coordinated campaigns over the Southern Ocean between  
38 Tasmania and Antarctica in January-February 2018 to evaluate the simulations of clouds  
39 and precipitation in nudged-meteorology simulations with the CAM6 and AM4 global  
40 climate models sampled at the times and locations of the observations. Fifteen  
41 SOCRATES research flights sampled cloud water content, cloud droplet number  
42 concentration, and particle size distributions in mixed-phase boundary-layer clouds at  
43 temperatures down to -25 C. The six-week CAPRICORN2 research cruise encountered  
44 all cloud regimes across the region. Data from vertically-pointing 94 GHz radars  
45 deployed was compared with radar-simulator output from both models. Satellite data was  
46 compared with simulated top-of-atmosphere (TOA) radiative fluxes.

47

48 Both models simulate observed cloud properties fairly well within the variability of  
49 observations. Cloud base and top in both models are generally biased low. CAM6  
50 overestimates cloud occurrence and optical thickness while cloud droplet number  
51 concentrations are biased low, leading to excessive TOA reflected shortwave radiation. In  
52 general, low clouds in CAM6 precipitate at the same frequency but are more  
53 homogeneous compared to observations. Deep clouds are better simulated but produce  
54 snow too frequently.

55

56 AM4 underestimates cloud occurrence but overestimates cloud optical thickness even  
57 more than CAM6, causing excessive outgoing longwave radiation fluxes but comparable  
58 reflected shortwave radiation. AM4 cloud droplet number concentrations match  
59 observations better than CAM6. Precipitating low and deep clouds in AM4 have too little  
60 snow. Further investigation of these microphysical biases is needed for both models.

61

## 62 **1. Introduction**

63 General circulation models (GCMs) are challenged by uncertainties and biases in  
64 the simulation of Southern Ocean clouds, aerosols, and precipitation, and these  
65 uncertainties affect simulated global cloud feedback on climate change. The clouds  
66 simulated by GCMs participating in the third and fifth Coupled Model Intercomparison  
67 Projects (CMIP3 & CMIP5; Meehl et al., 2005) mostly reflected too little sunlight back  
68 to space over the Southern Ocean (45°-65°S) (Trenberth and Fasullo, 2010; Ceppi et al.,  
69 2012; Williams et al., 2013). Bodas-Salcedo et al. (2014) and others identified  
70 insufficient low cloud cover and insufficient supercooled liquid water in the cold sector

71 of frontal cyclonic system as likely causes of this bias. Trenberth and Fasullo (2010)  
72 suggested that too little low cloud in the current climate might cause an underestimation  
73 of positive low cloud feedback on future climate change over this region. Models which  
74 glaciate mixed phase clouds at overly warm temperatures also have a spuriously negative  
75 high-latitude cloud optical depth feedback, driven by a simulated warming-induced  
76 transition from ice-dominated to liquid-dominated low clouds, while satellite  
77 observations suggest these clouds are already liquid-dominated (Cordon and Klein 2014;  
78 McCoy et al., 2016; Terai et al., 2016; Tan et al., 2016). Improved simulation of Southern  
79 Ocean clouds in climate models will help us to better simulate the radiative energy  
80 budget in the current climate and to make more reliable future projections of Earth's  
81 climate.

82         Several recent GCM sensitivity studies have shown that the SO cloud bias can be  
83 substantially reduced by inhibiting several uncertain stratiform and convective cloud  
84 microphysical processes that can glaciate mixed-phase Southern Ocean clouds (Kay et al.  
85 2016, Bodas-Salcedo et al., 2019; Gettelman et al., 2019). This may have led the Coupled  
86 Model Intercomparison Project phase 6 (CMIP6) versions of several GCMs (Eyring et  
87 al., 2016) with revised treatments of mixed-phase clouds to have more positive global  
88 cloud feedback than in their CMIP5 counterparts (Gettelman et al., 2019, Bodas-Salcedo  
89 et al., 2019, Zelinka et al., 2020).

90         Until recently, there were very few in-situ observations available to test and  
91 constrain such modeling choices. Satellite observations from active and passive sensors  
92 are an invaluable resource, but they have interpretational uncertainties that need to be  
93 anchored by in-situ measurements. An evaluation of the CMIP6 GCM simulations of SO

94 clouds and precipitation based on in-situ observations coordinated with collocated active  
95 remote sensing is a key step for future improvement of cloud representations in the  
96 models.

97         Motivated by this, two coordinated field studies were conducted over the sector of  
98 the Southern Ocean between Tasmania and the Antarctic sea ice edge in Jan.- Feb. 2018:  
99 1) a U. S. aircraft study based in Hobart, Tasmania, the Southern Ocean Clouds,  
100 Radiation, Aerosol Transport Experimental Study (SOCRATES), and 2) an Australian  
101 ship-based study, the second Clouds, Aerosols, Precipitation, Radiation, and atmospheric  
102 Composition Over the southeRn ocean field study (CAPRICORN2). These two studies  
103 used complementary sampling strategies. The research flights targeted weather regimes  
104 with low-lying clouds at altitudes below 4 km during daytime, providing detailed  
105 multivariate spatial cross-sections through complex cloud fields but no temporal  
106 continuity. The ship sampled all weather regimes and times of day, but its only in-situ  
107 measurements above the surface were twice-daily radiosondes. Both platforms had  
108 vertically-pointing cloud radar and lidar. The data from these two studies pair well  
109 because they test different aspects of GCM simulations.

110         In this paper, we use this data together with satellite measurements to characterize  
111 Southern Ocean cloud morphology, cloud and precipitation occurrence and frequency,  
112 cloud droplet number concentration ( $N_d$ ), hydrometeor size distribution, and shortwave  
113 (SW) and longwave (LW) radiative effects at the top of atmosphere (TOA). Radiosondes  
114 launched on the ship and dropsondes from the aircraft map out the troposphere relative  
115 humidity field. We use these uniquely comprehensive observations of cloud and radiative  
116 properties to evaluate the atmospheric components of two state-of-the-art CMIP6 GCMs.

117 The Community Atmosphere Model version 6 (CAM6, Bogenschutz et al., 2018) is the  
118 atmospheric component of version 2 of the Community Earth System Model (CESM2),  
119 developed by the National Center for Atmospheric Research (NCAR) and many other  
120 partners. The Atmosphere Model version 4 (AM4, Zhao et al. 2018) is part of the CM4  
121 climate model (Held et al. 2019) and ESM4 (Dunne et al. 2019) earth system model  
122 developed by the Geophysical Fluid Dynamics Laboratory (GFDL).

123 A centerpiece of our approach for comparing GCMs with observations is the use  
124 of nudged-meteorology simulations in which the GCM winds and temperature field are  
125 lightly nudged with a 24-hour timescale toward reanalysis, while other simulated fields  
126 (e. g. humidity, clouds, aerosols and precipitation) are not nudged and freely evolve.  
127 This allows us to focus on model errors in water processes that are probably derived from  
128 the local action of physical parameterizations rather than an incorrect synoptic  
129 environment.

130 The models are sampled along the same paths followed by the plane and the ship,  
131 so that every observation can be meaningfully compared with model output at the same  
132 simulated time and place, without need for compositing or other statistical averaging,  
133 similar to Wu et al. (2017) and Bretherton et al. (2019). The nudged-meteorology  
134 approach is particularly useful around the rapidly evolving storm systems of the SO.

135 Recently Gettelman et al. (2020) used SOCRATES and satellite measurements to  
136 look at cloud location, cloud phase, and boundary layer structure in CAM6 simulations,  
137 and evaluate the improvement of CAM6 simulations compared to CAM5 using monthly  
138 averaged satellite retrievals. Our paper complements Gettelman et al., (2020) by  
139 assessing cloud and precipitation occurrence and its radiative impacts from a more

140 statistical perspective, and combines unique CAPRICORN2 data and radar simulators for  
141 a comprehensive assessment.

142 The remainder of this paper is organized as follows. Section 2 describes our  
143 observations and models, including more detail on the nudged-meteorology approach  
144 taken here. Section 3 evaluates the representation of low cloud and precipitation in  
145 CAM6 and AM4 during the SOCRATES campaign, including cloud and precipitation  
146 occurrence and frequency, hydrometeor size distributions, cloud water content and cloud  
147 droplet number concentration. Section 4 discusses low and deep clouds in the models  
148 during the CAPRICORN2 campaign, using radar data and simulators and satellite-  
149 derived TOA radiative fluxes. Section 5 presents conclusions.

150

## 151 **2. Description of observations, models, and radar simulator**

### 152 **2.1. SOCRATES measurements**

153

154 During the SOCRATES campaign, 15 research flights of the U. S. National  
155 Science Foundation Gulfstream V (GV) research aircraft (EOL 2005) were conducted  
156 from Hobart, Tasmania (42°S, 147°E) out over the Southern Ocean between 15 January-  
157 24 February 2018. The GV aircraft flew roughly southward at its ferry altitude of 6 km to  
158 a southernmost waypoint, typically near 58-62°S, chosen to optimize sampling of cold-  
159 sector boundary-layer stratocumulus and cumulus. The GV then descended to conduct  
160 standardized sampling modules during the generally northbound return legs. Each 45-50  
161 minute module, spanning 400-500 km, was made up of 10-minute above-cloud, in-cloud,  
162 and below-cloud (150-200 m altitude) legs, and a sawtooth leg consisting of an ascent to

163 600 m above cloud top, a descent to 150 m above sea surface, and another ascent above  
164 the cloud top. A comprehensive suite of instrumentation for sampling mixed-phase cloud,  
165 aerosols, and turbulence was deployed ([https://www.eol.ucar.edu/content/socrates-](https://www.eol.ucar.edu/content/socrates-aircraft-payload)  
166 [aircraft-payload](https://www.eol.ucar.edu/content/socrates-aircraft-payload)), as well as a vertically-pointing cloud radar and lidar and dropsondes.

167 The primary in-situ instruments used in the current study are the Vertical-Cavity  
168 Surface-Emitting Laser (VCSEL; EOL 2008), the Cloud Droplet Probe (CDP), and the  
169 Two-Dimensional Stereo probe (2DS; Wu and McFarquhar, 2019). The VCSEL reported  
170 relative humidity (RH), derived as the ratio of measured water vapor concentration and  
171 saturated vapor pressure over liquid water at the ambient temperature (per Wexler's  
172 formula; Wexler 1976) at a 25 Hz temporal resolution. HARCO heated total air  
173 temperature sensors were used for measurement of temperature (T) every 25 Hz.

174 We use GV remote sensing measurements from the 94 GHz (W-band) HIAPER  
175 cloud radar (HCR; EOL 2014) and the high spectral resolution lidar (HSRL; EOL 2010),  
176 The radar and HSRL operated at a 2 Hz temporal resolution and could be manually  
177 switched to point up or down. The goal was generally to point toward the nearest clouds.  
178 Both instruments have a minimum range or 'dead zone' of 150-200 m from the plane, but  
179 this was rarely an issue unless the aircraft was flying within a thin cloud layer. Past its  
180 dead zone, the HSRL could detect essentially all clouds (with attenuation for thicker  
181 clouds), even when the aircraft was flying at its ferry altitude of 6 km. Thus, in this study  
182 the combination of the HSRL and the in-situ aircraft cloud probes were used to determine  
183 lower-tropospheric cloud occurrence.

184 The CDP measured liquid water content and cloud droplet size distribution from  
185 1-50  $\mu\text{m}$  at a sampling rate of 10 Hz. The 2DS provided hydrometeor images, from which

186 data processing software synthesized cloud and precipitation size distributions from 10-  
187 1028  $\mu\text{m}$  radius.

188

## 189 2.2: CAPRICORN2 measurements

190 The CAPRICORN2 cruise of Australia's Research Vessel (RV) *Investigator*  
191 spanned Jan. 10-Feb. 21, 2018. It was a sequel to earlier voyages in 20-29 March 2015,  
192 and March-April 2016 described in Protat et al., 2017 and Mace et al. 2018. We use radar  
193 reflectivity profiles collected by an onboard calibrated 95 GHz W-band vertically  
194 pointing cloud radar (see Mace et al. 2018 for more details). The radar reflectivity has  
195 been corrected for wet radome attenuation. We also use twice-daily radiosondes from the  
196 cruise.

197

## 198 2.3 Satellite measurements

199 To assess the GCM-simulated top-of-atmosphere (TOA) radiative fluxes, we use  
200 edition 4A of National Aeronautics and Space Administration (NASA) Clouds and the  
201 Earth's Radiant Energy System (CERES; Wielicki et al., 1996) synoptic (SYN) cloud and  
202 radiation products (Doelling et al., 2013; Rutan et al., 2015). We use the hourly TOA  
203 fluxes of reflected shortwave radiation (RSW) and outgoing longwave radiation (OLR).  
204 The CERES SYN data is available on a  $1^\circ \times 1^\circ$  grid  
205 (<https://ceres.larc.nasa.gov/products.php?product=SYN1deg>). We extract the nearest grid  
206 points to the contemporaneous aircraft and ship locations for comparison with models.

207

## 208 2.4 CAM6 model description

209 CAM6 was comprehensively described in Bogenschutz et al., (2018) and  
210 Gettelman et al. (2019). This section summarizes key features of CAM6 for this study.  
211 CAM6 implements the Cloud Layers Unified by Bi-normals (CLUBB, Golaz et al.  
212 (2002), Larson et al. (2002)) parameterization to replace the planetary boundary layer,  
213 shallow convection, and cloud macrophysical parameterization schemes used in CAM5.  
214 The unified CLUBB scheme bypasses the complexity of interactions between schemes to  
215 improve performance for the simulation of boundary layer clouds, especially of  
216 intermediate types of regimes such as the stratocumulus to cumulus transition  
217 (Bogenschutz et al., 2013; Guo et al., 2015). CAM6 retains the deep convection scheme  
218 of Zhang and McFarlane (1995) used in CAM4 and CAM5. The precipitation from the  
219 CLUBB and deep convection schemes is referred as large-scale (stratiform) and  
220 convective precipitation respectively. CLUBB diagnoses cloud fraction and cloud liquid  
221 water from a joint double-Gaussian probability density function (PDF). Ice and liquid  
222 cloud fractions in CLUBB are the same and are analytically diagnosed by integrating  
223 over saturated portions of the joint PDF (Guo et al., 2014). The total cloud fraction in  
224 CAM6 combines CLUBB and deep convective cloud cover fractions, and an ice cloud  
225 fraction assuming maximum overlap.

226 The CAM6 microphysics package incorporates a two-moment scheme for 4  
227 classes (liquid, ice, and large scale rain and snow) with updated ice nucleation  
228 parameterization, MG2 (Gettelman and Morrison, 2015). MG2 is coupled to a physically  
229 based mixed phase ice nucleation scheme (Hoose et al 2010) implemented in CAM6 with  
230 modifications for a PDF of contact angle by Wang et al (2014). MG2 accounts for  
231 preexisting ice during cirrus ice nucleation (Shi et al 2015).

232 Aerosols are predicted by a four-mode version of the Modal Aerosol Module  
233 (MAM4) (Liu et al., 2016), initialized based on climatological profiles in year 2000 from  
234 CMIP6 emissions inventory. The activation of aerosols into cloud droplets in CAM6 is  
235 diagnosed as a function of the modeled sub-grid scale updraft velocity and aerosol  
236 compositions and size distribution (Abdul-Razzak and Ghan 2000).

237 The CAM6 simulations in this paper are run with prescribed sea surface  
238 temperature. A Finite-Volume (FV) dynamical core of 0.9° longitude x 1.25° latitude  
239 resolution is used with 32 vertical levels and a model time step of 30 minutes. To  
240 facilitate model evaluation against observations, CAM6 was run in a nudged  
241 configuration (Lamarque, 2011) using the NASA Modern-Era Retrospective analysis for  
242 Research and Applications version 2 (MERRA-2; Rienecker et al., 2011; Molod et al.,  
243 2015) horizontal winds, temperature, and monthly mean sea surface temperature (SST)  
244 with a relaxation timescale of 24 hours. MERRA-2 nudging fields are interpolated to the  
245 CAM6 vertical levels before nudging. The CAM6 simulation is performed starting on  
246 January 1<sup>st</sup> 2017, to ensure proper spin-up of aerosol and land-surface fields well before  
247 any observational comparisons. Model outputs along the tracks of the aircraft and ship  
248 (specifically, from the nearest model grid points to the current ship and aircraft locations)  
249 are calculated in-line and output at time steps of 1 minute and 10 minutes respectively.

250

## 251 **2.5 AM4 model description**

252 AM4 was comprehensively described by Zhao et al. (2018). Here we summarize  
253 those physical parameterizations from the model that are particularly relevant to its  
254 simulation of Southern Ocean clouds and aerosols. AM4 uses a double plume shallow

255 convection scheme adapted from Bretherton et al., (2004), and a deep convection scheme  
256 based on a cloud work function relaxation closure (Zhao et al., 2018). The macrophysical  
257 scheme of large-scale clouds in AM4 follows Tiedtke (1993). Cloud water content and  
258 fractional cloud cover are described prognostically by large-scale budget equations. The  
259 increase in cloud cover is determined by the fraction of the cloud-free area exceeding  
260 saturation. AM4 implements a one-moment microphysics scheme for liquid water  
261 following Rotstayn (1977) and Rotstayn et al., (2000) with an inclusion of a prognostic  
262 scheme for cloud droplet number concentration (Ming et al., 2007), as in AM3. A rain  
263 profile is diagnosed at each time from the cloud properties (Rotstayn et al., 1997).

264       Ice is predicted from water vapor diffusion at the expense of liquid water (the  
265 Wegener-Bergeron-Findeisen process) and homogeneous freezing of liquid water at  
266 temperatures colder than  $-40^{\circ}\text{C}$ . Ice melts to form liquid water at temperatures warmer  
267 than  $0^{\circ}\text{C}$ . In AM4, there is no distinction between falling ice, snowflakes and graupel. All  
268 forms of atmospheric ice are represented by a single variable. The ice particles fall with a  
269 mass-weighted mean velocity calculated assuming fall speed is proportional to the 0.16  
270 power of particle diameter. Falling ice particles are approximated by a negative  
271 exponential distribution with effective radius determined by temperature that ranges from  
272  $15 - 100 \mu\text{m}$  (Donner et al., 1997).

273       Aerosols in AM4 are predicted based on climatological sources in year 2016 from  
274 the CMIP6 emissions inventory; only the mass is prognosed for each aerosol type with a  
275 fixed assumed size distribution (Zhao et al., 2018). The activation of aerosols into  
276 droplets uses the parameterization of Ming et al. (2006).

277 AM4 uses the GFDL Finite-Volume Cubed-Sphere dynamical core (FV<sup>3</sup>; Harris  
278 and Lin, 2013; Putman and Lin, 2007) with a grid of approximately 100 km horizontal  
279 resolution and 33 vertical levels. For the simulations presented here, AM4 was run in a  
280 nudged configuration (Jeuken et al., 1996) similar to that used for CAM6, with the same  
281 24 hour nudging timescale, but instead nudged to the fifth generation of the European  
282 Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis of the  
283 global climate (ERA5; Hersbach and Dee, 2016) horizontal winds, temperature, and  
284 surface pressure with a relaxation time of 24 hours. Like CAM6, the AM4 simulation  
285 starts on January 1<sup>st</sup> 2017. Data is output every 3 hours for radiation fields and 1 hour for  
286 other quantities. The nearest model grid points to the ship and aircraft locations were  
287 extracted from the AM4 simulations by linearly interpolating to the observation point for  
288 comparison with observations and CAM6.

289

## 290 **2.7 COSP radar simulator**

291

292 Within each grid column, the profiles of cloud and precipitation are converted to  
293 profiles of synthetic radar reflectivity using implementations of the Cloud Feedback  
294 Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP; Bodas-  
295 Salcedo et al., 2011) in the two GCMs. CAM6 and AM4 use COSP version 2.1 and 1.4.1  
296 respectively (Bodas-Salcedo et al., 2011; Swales et al., 2018), but there is no crucial  
297 scientific difference between COSP versions. In this study, we focus on use of the  
298 CloudSat simulator within COSP. It provides synthetic radar reflectivity at a frequency of  
299 94 GHz and can be compared with the observed W-band reflectivity.

300           The implementation of COSP in a GCM usually makes some additional model-  
301 specific assumptions that are not part of the GCM, are not necessarily well documented,  
302 and which may impact the synthetic radar reflectivity. For example, the hydrometeor size  
303 distribution assumptions can be slightly different between COSP and the parent GCM  
304 microphysics scheme. In the CAM6 COSP, all hydrometeors are described with modified  
305 gamma distributions. In the CAM6 microphysics scheme, cloud drops are described with  
306 a gamma distribution while ice, rain, and snow are assumed to have exponential  
307 distributions (gamma with  $m=0$ ).

308           The AM4 microphysics scheme has a single ice category that includes both cloud ice and  
309 snow and has an aggregate fall speed. In this sense, snow is simply falling ice. AM4  
310 treats the total ice and snow concentration as cloud ice in COSP, which is assigned to  
311 have the temperature-determined effective radii of cloud ice particles in AM4.  
312 Furthermore, the clear-sky ice flux (flux of ice outside of cloud entering the unsaturated  
313 portion of the grid box from above) is used for snow in COSP with effective radii  
314 computed internally in COSP. Snow inside clouds is not accounted for explicitly. The  
315 impact on the synthetic radar reflectivity of these differences in the assumptions made  
316 between COSP and the GCM microphysics scheme is discussed in Appendix B.

317           The COSP interface varies between host models. CAM6 uses COSP's default -  
318 column generator to produce 10 homogenous sub-columns, while AM4 treats the sub-  
319 grid cloud and precipitation fields from the radiation scheme as the COSP sub-columns,  
320 rather than using the default COSP sub-column generator. We observed little difference  
321 between the sub-columns. The insufficient sub-column variability in COSP's default sub-

322 column generator may lead to overestimated radar reflectivity and probability of  
323 precipitation compared to the satellite observations (Song et al., 2018).

324

325

### 326 **3. Low clouds and precipitation in CAM6 and AM4 simulations during** 327 **SOCRATES**

328

329 In this section, we will use in-situ and remote sensing observations from SOCRATES  
330 to evaluate the macrophysical and microphysical properties of clouds and precipitation in  
331 CAM6 and AM4. SOCRATES sampling focused on low clouds with cloud top height  
332 lower than 4 km and little or no precipitation falling from any overlying clouds through  
333 the 4 km level. We select RF09 (a case of cumulus rising into stratocumulus) and RF12  
334 (a stratocumulus case) as two examples to demonstrate single-flight comparisons of  
335 observations and GCM simulations of shallow cumulus and stratocumulus regions,  
336 followed by cloud-related statistics across the whole campaign.

337

#### 338 **3.1 RF09 temperature, relative humidity, cloud and precipitation comparisons**

339 Fig. 1 shows time-height plots of T, RH, in-cloud cloud water content (CWC),  
340 and precipitating particle number density ( $N_{Large}$ , described below) along Flight RF09  
341 (inside the black channel) overlying the corresponding fields simulated by CAM6 and  
342 AM4 respectively. The microphysical fields are only plotted over the 0-4 km altitude  
343 range to highlight low clouds and their environment, while the thermodynamic fields are

344 plotted from 0-8 km altitude to encompass the ferry leg and provide synoptic-scale  
345 context.

346 RF09 targeted an extensive deck of cold, low-level cloud in the cold sector of a  
347 mid-latitude cyclone south and east of Tasmania. Two sampling modules were completed  
348 in the cold sector regions south of 50°S. As seen in Figs. 1a and 1e, the boundary-layer  
349 cloud tops, at a height of 2.5 km, have a cloud top temperature near -15°C. The  
350 temperature in the two nudged GCM simulations agrees with the in-situ observations to  
351 within 1-2 C. Since temperature is a nudged field, this indicates that the nudged-  
352 meteorology approach is working as hoped.

353 Relative humidity (Fig. 1b) is important for producing clouds. It is a more  
354 challenging test for the nudged GCM simulations, since their humidity fields are not  
355 constrained with reanalysis data. For both observations and models, the RH in this paper  
356 is computed based on liquid saturation. In RF09, the high-RH boundary layer is capped  
357 by dry, low-RH, subsiding air above 2.5 km. The free-tropospheric RH is fairly well  
358 simulated by both models. Inside the boundary layer, the observed RH is horizontally  
359 variable, and is relatively low in the ascent portion of a cloud-free sawtooth near 57°S in  
360 the return leg of RF09. This is suggestive of shallow cumulus rising into a broken  
361 stratocumulus layer, a common cold-sector cloud type. As seen in Figs. 1b and 1c, both  
362 models capture the boundary layer depth qualitatively well except that they underestimate  
363 RH at the top of the boundary layer. The boundary-layer RH in CAM6 is comparable to  
364 observations (Fig. 1b), but the AM4 boundary layer is drier than observed (Fig. 1c).

365 Figs. 1c and 1g show the observed and modeled in-cloud water content (CWC)  
366 during RF09. This is an even more challenging comparison for the models because it

367 requires the models to have both accurate cloud placement and cloud microphysics. The  
368 observed CWC is taken from the GV CDP and is plotted when its value exceeds  $0.01 \text{ g}$   
369  $\text{m}^{-3}$ . CWC less than  $0.01 \text{ g m}^{-3}$  is masked in gray. For the GCMs, the cloud-containing  
370 grid cells are distinguished from clear-sky grid cells by having nonzero cloud water  
371 mixing ratio and the in-cloud water content is calculated by dividing grid-mean cloud  
372 water content by simulated cloud fraction. To be consistent with observations, the GCM  
373 CWC is plotted when its value exceeds  $0.01 \text{ g m}^{-3}$ .

374 To shed light on the representation of precipitation in the GCMs, we compute in-  
375 cloud NLarge (Fig. 1d). NLarge is computed from 2DS particle size distributions (PSD)  
376 as the concentration of large precipitating particles with radius greater than 100 microns.  
377 The observed NLarge is compared against the CAM6 counterpart along the flight track  
378 computed in the same way as in observations based on the model PSD of fraction mean  
379 cloud and precipitation. The 'fraction-mean' cloud and precipitation are calculated by  
380 dividing grid-mean cloud and precipitation quantities by simulated cloud and  
381 precipitation fraction respectively. The CAM6 precipitation fraction is set to be the same  
382 as the cloud fraction in each cloud-containing grid cell and to the cloud fraction of the  
383 lowest cloud-containing grid cell below cloud. Because precipitation in AM4 is treated  
384 diagnostically, NLarge is not computed by AM4.

385 The RF09 sawtooth legs sampled a broken cloud field with intermittent CWC  
386 (Fig. 1c). CAM6 generally underestimates its cloud water content (Fig. 1c) but  
387 overestimates NLarge (Fig. 1d). The CWC in AM4 in RF09 agrees better with  
388 observations than CAM6, but the AM4 clouds have lower cloud base heights compared  
389 to observations, a bias seen in many cases during the SOCRATES campaign.

390

### 391 **3.2 RF12 temperature, relative humidity, cloud and precipitation comparisons**

392 Fig. 2 compares observations and simulations for an extensive stratocumulus case  
393 sampled during RF12 in two modules south of 55°S. The stratocumulus deck topped a  
394 fairly well-mixed 1500 m deep boundary layer, with a cloud top temperature around -9°C  
395 capped by a 5°C temperature inversion. The cloud deck was in the cold sector of a weak  
396 cyclone. Figs. 2a and 2b confirm that the temperature in the nudged models is consistent  
397 with the observations for RF12, like in RF09. Both CAM6 and AM4 clearly show low  
398 RH at the top of the boundary layer, suggesting biased low boundary layers in these  
399 GCMs. As one might expect, the CAM6 CWC and NLarge in the comparatively  
400 horizontally homogeneous stratocumulus decks of RF12 agree better with observations  
401 than these quantities in the more heterogeneous cumulus regions of RF09. However,  
402 CAM6 also tends to miss the light precipitation and spatially intermittent snow that  
403 formed in the thicker centers of mesoscale closed cells during RF12 (e.g., -58°N and -  
404 56°N in the return flight in Fig. 2d). CAM6 and especially AM4 simulate a cloud base  
405 height that is too low compared to the approximately 1 km base observed in RF12 (Figs.  
406 2c and 2g). In AM4, the simulated clouds extend down to the ground level.

407

### 408 **3.3 Statistical all-flight comparisons of temperature, relative humidity, cloud and** 409 **precipitation**

410 In order to test the accuracy of the large-scale meteorology in the GCMs, the root  
411 mean square (RMS) error was calculated between the observations and GCMs. Across all  
412 campaign flights, observed temperature and humidity along the flight track were

413 averaged over 50 m bins in altitude during each 2 minute time interval. The two nudged  
414 GCMs were similarly sampled. CAM6 and AM4 had RMS temperature errors of 1.3 K  
415 and 1.4 K, remarkably small considering the remote sampling region and large synoptic  
416 variability. This is mostly a testament to the accuracy of the reanalysis to which the  
417 GCMs were being nudged (which match the observations within even smaller RMS  
418 errors of less than 1 K). However, it also shows both GCMs are very good short-term  
419 weather forecast models that are able to retain this level of accuracy for at least a day (the  
420 nudging timescale).

421 Humidity is highly variable and was not nudged, so it is a much more challenging  
422 comparison for the models. We use RH as a measure of humidity, since it has  
423 comparable variability across RMS errors across the whole range of sampled heights.  
424 Across all flight samples, the RMS error of RH is 23% and 22% for CAM6 and AM4  
425 respectively. For comparison, the ERA5 and MERRA-2 reanalysis had slightly smaller  
426 RMS RH errors of 17% and 19%. Such errors are large enough to affect the existence  
427 and placement of cloud layers, even in a GCM with perfect microphysics.

428 Cloud placement errors reduce the value of a point-by-point comparison of GCM  
429 vs. observed cloud properties. It is more illuminating to make a statistical comparison of  
430 mean biases in GCM vs the observed CWC at the same overall region, altitude range, and  
431 time. We bin the observed and simulated CWC for the 15 SOCRATES flights into boxes  
432 of 500 m in altitude and 25 minutes (equivalent to 210 km at a typical flight speed of 140  
433  $\text{m s}^{-1}$ ) in time along the flight track. This binning box is chosen to be big enough to  
434 reduce sampling noise but small enough to still represent the local CWC. Boxes in which  
435 the binned average CWC  $< 0.01 \text{ g m}^{-3}$  for either the models or the observations are

436 excluded from the statistics. Boxes with less than ten observed samples are screened out.  
437 This leaves 133 binned samples, most of which are in altitudes below 3 km. Fig. 3  
438 presents the bin-mean and range of CWC over all time bins within each altitude band.  
439 The model and observed CWC interquartile ranges generally agree with each other  
440 between 1.5~2 km (although with large spread). CWC is clearly overestimated, especially  
441 by AM4, below 1 km, an indication that the simulated cloud base is systematically too  
442 low as in the RF09 and RF12 examples. On the other hand, in-cloud CWC for both  
443 GCMs, especially AM4, is biased low above 2.5 km compared to observations,  
444 suggesting that the GCM clouds tend to have a slightly lower cloud top height. This is  
445 consistent with the low RH bias at the top of the boundary layer seen for RF12 in Figs. 2b  
446 (CAM6) and more prominently in Fig. 2f (AM4).

447

### 448 **3.4 Low cloud occurrence**

449 Occurrence of low clouds with tops below 4 km in CAM6 and AM4 columns cannot  
450 be evaluated using the in-situ observations, since they targeted cloud layers. Instead, it is  
451 evaluated in this section using a column cloud fraction based on combining a HSRL  
452 backscatter threshold to detect cloud above or below the aircraft and the GV CDP liquid  
453 water content to detect cloud at the aircraft level which may not extend outside the 150 m  
454 lidar dead zone, or which may attenuate the lidar beam before it reaches the cloud edges.  
455 Within a lidar sampling time of 0.5 s, low cloud is flagged if any of the 10 Hz CDP liquid  
456 water content measurements exceeds  $10^{-4}$  kg m<sup>-3</sup> below 4 km, or if the maximum HSRL  
457 backscatter below 4 km altitude exceeds a threshold of  $3 \times 10^{-5}$  m<sup>-1</sup> sr<sup>-1</sup>. This backscatter

458 threshold effectively separates cloud echoes from those of aerosols, as documented in  
459 Appendix A.

460 Examples of the lidar backscatter for RF09 and RF12 are shown in Figs. 4a and 4e,  
461 where cloud boundaries (i.e., cloud tops when the aircraft was above and cloud bases  
462 when below) are well captured by HSRL as seen from the strong lidar backscatter near 1  
463 to 2 km. The observed upper cloud boundaries (cloud tops) are slightly higher than those  
464 implied by the GCM cloud fraction maps.

465 We define the observed low cloud fraction as the fraction of low cloud flags during  
466 every 10 minutes (equivalent to  $\sim 1$  degree at a typical flight speed of  $200 \text{ m s}^{-1}$ ). We  
467 compare this with the corresponding low cloud fraction averaged over the same time  
468 periods when there is observational data in CAM6 and AM4 (e.g., Figs., 4b, c, f, and g).  
469 The low cloud fraction for each GCM is computed following that GCM's vertical cloud  
470 overlap assumptions (maximum-random overlap for CAM6 and exponentially decaying  
471 overlap for AM2 with a length scale of 2 km (Zhao et al., 2018)). The regions outside of  
472 the HSRL view zone (i.e., regions above/below the aircraft when the HSRL pointed  
473 down/up) are masked out before computing GCM low cloud fraction (grey shading in  
474 Fig. 4).

475 The low cloud fraction comparisons for RF09 and RF12 are shown in Figs. 4d and  
476 4h. As suggested by the lidar backscatter profiles in Fig. 4a and 4e, the observed low  
477 cloud fraction in the cumulus regions in RF09 is smaller than that in the stratocumulus  
478 regions in RF12. In both flights, CAM6 typically simulates a low cloud fraction that is  
479 too large, whereas that in AM4 is too small.

480 Similar low cloud fraction biases are present across the 15 SOCRATES flights. Fig. 5a  
481 shows an all-flight histogram of 10-minute average low cloud fraction. Low clouds,  
482 either alone or co-occurring with cloud layers aloft, are observed in 96% of the 10-minute  
483 intervals during SOCRATES. About half of the intervals have a low cloud fraction  
484 greater than 80%. Only ~10% of the intervals have a low cloud fraction less than 20%. In  
485 CAM6, intervals of nearly complete low cloud cover (greater than 90%) occur 60% of  
486 the time vs. ~30% of the time in AM4 and 45% in the observations. Over half of the  
487 intervals including low clouds in AM4 are characterized by a low cloud fraction smaller  
488 than 50%, about twice as frequent as CAM6 and observations.

489 Another way to present this data is by binning the 10-minute intervals by the  
490 observed low cloud fraction, and testing how well the models replicate the low cloud  
491 fraction within each bin (Fig. 5b). Ideally, a model would lie on the 1:1 line with no  
492 scatter about the observations in this box-whisker plot, but from our other comparisons  
493 we expect both large scatter (a large interquartile range of simulated cloud fraction for a  
494 given observed cloud fraction) and bias. Indeed, the scatter is large, and the interquartile  
495 ranges show that in most bins, about 75% of the CAM6 samples lie above the observed  
496 cloud fraction, while about 60% of the AM4 samples lie below the observed cloud  
497 fraction. One exception for AM4 is that it produces too much cloud when the observed  
498 cloud fraction is less than 10%. This could be due to geographical misplacement of  
499 scattered cloud rather than parameterization biases given its agreement with observations  
500 for the 10-20% low cloud fraction bin. In summary, CAM6 overestimates and AM4  
501 underestimates low cloud fraction in the cold-sector low cloud regimes sampled by  
502 SOCRATES.

503

### 504 **3.5 TOA upwelling SW and OLR**

505 Biases in CWC and low cloud fraction contributes to radiative biases in the GCMs. A  
506 conventional way to evaluate the impact of cloud on radiation is to compute cloud  
507 radiative forcing, defined as the difference of net downward radiative fluxes at TOA with  
508 and without cloud. However, since the retrieval of clear-sky radiation from satellite  
509 observations inevitably involves uncertainty, in this study we instead compare observed  
510 and simulated TOA reflected shortwave and outgoing longwave radiative fluxes as more  
511 reliably observed proxies for cloud effects on radiation. We recognize that they may also  
512 incorporate biases not related to cloud, e. g. in humidity or surface properties. The  
513 radiative flux estimates are matched to the same locations and times as the low cloud  
514 fraction estimates.

515 Fig. 6 shows the TOA RSW and OLR fluxes along the flight tracks from CERES  
516 SYN (Section 2.3) and from the two models, binned by observed low cloud fraction.  
517 Consistent with the overestimated cloud fraction in CAM6, the RSW in CAM6 is biased  
518 high for all bins of observed low cloud fraction. This high bias remains significant even  
519 when the observed low cloud fraction is 90-100%, suggesting that the low clouds in  
520 CAM6 are not only too frequent, but also too bright. As a result, the average RSW in  
521 CAM6 over the entire SOCRATES field campaign is about 20% higher than observed.  
522 The overestimate of low cloud cover in CAM6 also leads to underestimated OLR in bins  
523 with 50% or less observed low cloud cover. Since the CAM6 cloud tops are at altitudes  
524 comparable to observed, although slightly low-biased, they appear not to have large  
525 cloud-top temperature biases. Thus, when the observed and CAM6 cloud fractions are

526 near to 100%, the average OLR of CAM6 is similar to observed. The radiation bias of  
527 CAM6 ('too frequent, too bright') is consistent with the climatological cloud radiative  
528 effect shown in Gettelman et al. (2020).

529 In contrast, the underestimated low cloud fraction in AM4 allows for more OLR  
530 originating from the sea surface to escape to space, contributing to a sizable high OLR  
531 bias in all cloud fraction bins. Surprisingly, the AM4 TOA upwelling SW is comparable  
532 to observations in all observed cloud fraction bins. This implies the clouds are optically  
533 thicker than observed, i. e. AM4 has a 'too few, too bright' bias for SO low clouds, which  
534 is common in CMIP5 models (Nam et al., 2012; Engstrom et al., 2015).

535

### 536 **3.6 Microphysics in precipitating and non-precipitating low clouds**

537 We now investigate some underlying model-observation discrepancies in  
538 microphysics that may contribute to the radiation biases in models associated with  
539 Southern Ocean low clouds.

540 We quantify the occurrence of precipitating and non-precipitating low clouds in  
541 observations and CAM6 along the flight track sorted by ambient temperature (Fig. 7a).  
542 An observed or CAM6 low cloud is classified as precipitating if  $N_{Large}$  (defined in  
543 Section 3.1 as the concentration of cloud particles with radius bigger than  $100 \mu\text{m}$ ; recall  
544 also that this cannot be computed for the simpler AM4 microphysics) is greater than  $1 \times$   
545  $10^{-4} \text{ m}^{-3}$  in observations or CAM6 simulations. The occurrence is computed in cloud  
546 regions where CDP CWC exceeds  $0.01 \text{ g m}^{-3}$ . Eighty-five percent of the SOCRATES  
547 samples were collected in cold clouds (at temperatures below freezing), of which only  
548 ~10% were precipitating. This is partly because the GV intentionally avoided long flight

549 legs in drizzling supercooled clouds for safety. Repeating the analysis based on the  
550 nearest CAM6 grid cells along all 15 SOCRATES flight tracks (Fig. 7b), we find that the  
551 CAM6 clouds span a generally similar temperature with comparable precipitation  
552 occurrence, although precipitation occurrence in CAM6 clouds does not agree that well  
553 with observed clouds during individual flights (e.g., precipitation is overestimated in  
554 RF09 but underestimated in RF12 in CAM6; Figs. 1d and 2d). The imperfect match  
555 during individual flights might be because the deficient representation of the cloud  
556 intermittency in CAM6.

557 We compared the hydrometeor size distributions observed from the CDP and 2DS  
558 averaged over the nonprecipitating and precipitating clouds with those inferred along the  
559 flight tracks from CAM6 (Fig. 8), summed over cloud, rain, ice and snow. As seen in Fig.  
560 8a, nonprecipitating clouds display a unimodal distribution with a peak around 10  $\mu\text{m}$   
561 radius. This unimodal distribution is well represented in CAM6 and is dominated by  
562 liquid. CAM6 underestimates the number of cloud droplets with radii less than 20  $\mu\text{m}$ ,  
563 which dominate the overall cloud droplet number concentration. This bias is larger for the  
564 precipitating clouds (Fig. 8b).

565 By definition, the observed number of particles with radius  $> 50 \mu\text{m}$  is larger for  
566 precipitating clouds, leading to a shoulder in the observed droplet size distribution seen in  
567 Fig. 8b. The CAM6 simulations have a comparable increase in rain (blue dash) at 50-300  
568  $\mu\text{m}$  radii and in snow (red dash) at radii exceeding 300  $\mu\text{m}$ , suggesting that there is  
569 slightly more snow on average in CAM6 than in observations. The model PSDs should  
570 not be expected to agree perfectly well with observations on the large-radius tail, given a  
571 simple bulk two-moment scheme in CAM6. Note that the PSD in this study is computed

572 from in-cloud legs defined as  $CWC > 0.01 \text{ g m}^{-3}$ . CAM6 is found to have more rain than  
573 observations if a less strict in-cloud threshold is used (Gettelman et al., 2020).

574

### 575 **3.7 Phase partitioning**

576 The supercooled boundary-layer clouds sampled by the GV at temperatures of -5  
577 to -25°C were a mix of small liquid drops that dominate the cloud optical depth and  
578 (when precipitating) larger ice and snow particles. This conclusion is based on several  
579 complementary lines of evidence.

580 We visually inspected representative images from the 2DS and the PHIPS HALO  
581 (Schnaiter, 2018), a new imaging instrument deployed on the GV for SOCRATES that is  
582 optimized to detect ice particles with radii between 20-300  $\mu\text{m}$  and liquid drops with radii  
583 of 60-300  $\mu\text{m}$  (Abdelmotalieb et al., 2016; Schnaiter et al., 2018). These images suggest  
584 that in the precipitating boundary-layer clouds sampled by the GV at temperatures of -5  
585 to -25°C, most of the larger particles (radius  $> 100 \mu\text{m}$ ) are aspherical frozen  
586 hydrometeors.

587 The SOCRATES 2DS data have insufficient spatial resolution to clearly  
588 discriminate the phase of small particles with radii less than 100  $\mu\text{m}$ . We instead used a  
589 comparison between the liquid water content inferred from the CDP and from a CSIRO  
590 (The Commonwealth Scientific and Industrial Research Organization) King hotwire  
591 probe to test for the presence of small ice particles of radius less than 25  $\mu\text{m}$ , the size  
592 range dominating the cloud droplet number concentration and thus optical depth. Such  
593 particles would be detected by the CDP but the data processing algorithm would treat  
594 them as liquid water droplets, which introduces a high bias in CDP-inferred cloud water

595 content due to their lower density. Small ice particles should affect the CSIRO King  
596 probe's LWC measurement rather differently. For instance, ice might partly bounce off  
597 the hot wire causing the King probe to underestimate the cloud ice contribution to the  
598 cloud water content. Hence a comparison of the LWC inferred from the two instruments  
599 can test the presence of cloud ice. Fig. 9 shows a two-dimensional histogram of the two  
600 LWC measurements over all SOCRATES low cloud sampling at temperatures -5 to -  
601 25°C, presented as a two-dimensional histogram. The strong concentration of data along  
602 the 1:1 line is evidence that small particles (radius < 25 um) are predominantly  
603 supercooled liquid droplets.

604 Mace et al., 2018 reports that the light scattering from supercooled Southern  
605 Ocean boundary layer stratocumulus clouds mostly comes from liquid droplets, based on  
606 an analysis of ship-borne lidar depolarization ratios during CAPRICORN. Our visual  
607 inspection of plots of HSRL depolarization ratios from boundary-layer cloud tops  
608 observed during SOCRATES supports this conclusion.

609 The hydrometeor PSDs in CAM6 (Fig. 8) are also dominated by supercooled  
610 liquid droplets at small sizes.

611

### 612 **3.8 Cloud droplet number concentration ( $N_d$ )**

613 We compare observed in-cloud  $N_d$ , computed as the summation of cloud droplets  
614 measured by the CDP when the CDP CWC > 0.01 g m<sup>-3</sup>, with the GCM-simulated in-  
615 cloud  $N_d$ . Fig. 10 shows the RF09 and RF12 examples. AM4  $N_d$  is comparable to  
616 observations, but CAM6 significantly underestimates  $N_d$ .

617 These flights are representative of SOCRATES as a whole. Fig. 11 shows  
618 interquartile range boxes of observed and GCM in-cloud  $N_d$  measured across all 15  
619 SOCRATES flights and binned similarly to the in-cloud CWC described in Section 3.1.  
620 Points where binned average  $N_d < 1 \text{ cm}^{-3}$  for either the models or the observations are  
621 excluded from the statistics. Fig. 11 shows that the observed  $N_d$  clusters around 25-150  
622  $\text{cm}^{-3}$  with the highest  $N_d (> 100 \text{ cm}^{-3})$  occurring mostly near 0.5-1.5 km. CAM6 shows a  
623 low bias in  $N_d$  above 500 m which amplifies with height. AM4 simulates more high  $N_d$   
624 outliers than observed for clouds above 2 km, and does not simulate the relatively  
625 uncommon occurrences of observed  $N_d$  lower than  $40 \text{ cm}^{-3}$ . On average, however, AM4  
626 produces a mean  $N_d$  at all altitudes much closer to observations than CAM6.

627 CAM6's low  $N_d$  bias could be due to insufficient CCN production or too small a  
628 fraction of aerosol activated in the model. McCoy et al., (2020b, in prep) finds that  
629 CAM6 simulates CCN concentrations fairly well during SOCRATES with no significant  
630 low bias. We find that there is no significant statistical bias in precipitation scavenging of  
631 CCN in CAM6 when all cases are considered. Atlas et al. (2020) finds CAM6 simulates  
632 too little cloud-layer turbulence in stable and neutral boundary layers, which could lead to  
633 an under activation of CCN. However, CAM6 *also* underestimates  $N_d$  in unstable  
634 boundary layers for which its simulated turbulence is on average consistent with  
635 observations. This suggest that there may be multiple competing biases in the model.  
636 Disentangling these compounding influences will be necessary to understand the cause of  
637  $N_d$  bias in CAM6 and should be the topic of future investigations.

638

639 **4. Clouds and precipitation in CAM6 and AM4 simulations during the**  
640 **CAPRICORN2 campaign**

641 The upward-pointing 94 GHz shipborne radar deployed on the *R/V Investigator*  
642 during CAPRICORN2 sampled whatever clouds were overhead, including many periods  
643 of deep clouds with cloud tops above 4 km that were not targeted in SOCRATES. We use  
644 this unique radar dataset to evaluate the representation of both deep and low clouds in the  
645 GCMs.

646

647 **4.1 Relative humidity, cloud morphology, and TOA radiative fluxes**

648 We use the 1-15 February, 2018 period of the CAPRICORN2 campaign to  
649 illustrate typical model biases. Fig.12a shows a time-height section of radar reflectivity  
650 during this period. Low clouds with cloud tops below 4 km were regularly observed  
651 while deep cloud layers reaching above 6 km were also frequent. The deep clouds are  
652 often associated with significant precipitation indicated by strong reflectivity ( $>0$  dBZ)  
653 near the surface, which also often attenuates the W-band radar echo below detectability  
654 above 6 km. The precipitation from the thin low clouds is much weaker. As one would  
655 expect, the cloudy, precipitating regions are collocated with high relative humidity in a  
656 time-height section created from the twice-daily ship-launched radiosondes (Fig. 12b).  
657 The RH (computed based on liquid saturation) is shown in Fig. 10c and 10d for CAM6  
658 and AM4. Both models qualitatively reproduce the RH profiles for low cloud regimes  
659 sampled along the ship track. CAM6 slightly overestimates the observed RH in regions of  
660 deep cloud while AM4 substantially underestimates RH in those regions and also

661 simulates a shallower cloudy boundary layer than observed, as we also saw in the  
662 SOCRATES airborne data (Figs. 1b-f, 2b-f).

663 Fig. 13 compares TOA RSW (a) and OLR (b) from CERES SYN observations  
664 with the two models for the same period during CAPRICORN2. The deep clouds in  
665 CAM6 tend to reflect more shortwave radiation (are 'brighter') than observed, leading to  
666 a 10% high bias in the mean reflected SW over the whole period. The CAM6 OLR has a  
667 time-mean comparable to the observations but has a low bias in the deep cloud regions  
668 (e.g., Feb. 1-3, 11). In AM4 the RSW is comparable to CERES with intermittent high  
669 biases, while the OLR is typically slightly high. Overall, these biases are similar to those  
670 discussed in Section 3.3 for low clouds observed in SOCRATES. They imply that deep  
671 clouds, like low clouds, are in general too bright in both CAM6 and AM4, and are too  
672 frequent in CAM6 but too broken in AM4.

673

#### 674 **4.2 Comparison of observed and simulated radar reflectivities**

675

676 Fig. 14 shows reflectivities from the CAM6 and AM4 COSP simulators for the  
677 CAPRICORN2 campaign. For this study, CAM6 COSP provided reflectivity with and  
678 without hydrometeor and gas attenuation as viewed from the ground (Figs. 14a and 14b),  
679 while AM4 COSP only output attenuated reflectivity as viewed from space (Fig. 14c). As  
680 seen by comparing Figs. 14a and 14b, the inclusion of attenuation can reduce the  
681 reflectivity by several dB for deep precipitating clouds, but it has no significant impact on  
682 cloud morphology and low cloud reflectivity. Since AM4 COSP reflectivity is  
683 significantly weaker than that of CAM6 COSP (Fig. 14c), the hydrometeor attenuation is  
684 of only minor importance. As such, we expect the space-based attenuated reflectivity of

685 AM4 COSP to be qualitatively comparable to its ground-based counterpart. In the rest of  
686 the study, unless otherwise mentioned, we will compare attenuated CAM6 and AM4  
687 COSP reflectivity with observations.

688 CAM6 COSP reflectivity (Fig. 14b) agrees fairly well with the ship-observed  
689 reflectivity (Figs. 12a), but has longer and less interrupted periods of deep cloud  
690 occurrence (e.g., Feb. 1-3; Feb. 11-13). The AM4 COSP reflectivity is significantly too  
691 weak in the deep clouds, indicating underestimation of snow (Fig. 14c), for reasons to be  
692 discussed in Section 4.3. An abrupt change in reflectivity occurs at the freezing level at 1-  
693 2 km, below which the AM4 COSP reflectivity matches the observations better.

694

695

### 696 **4.3 Low and deep clouds**

697 For a quantitative statistical comparison of observed and modeled reflectivity, we  
698 construct Contoured Frequency by Altitude Diagrams (CFADs, Yuter and Houze, 1995)  
699 of observed and COSP reflectivity along the entire ship track during the CAPRICORN2  
700 campaign (Fig. 15). The joint histograms are created for every 2 hours with a 100 m  
701 vertical resolution and 2 dBZ increments from -40 dBZ to 10 dBZ in the horizontal, then  
702 conditionally averaged over the desired cloud regimes. Unlike in some studies of deep  
703 convection (e.g., Houze et al., 2007), our CFADs are not normalized to exclude regions  
704 with no detectable reflectivity.

705 The CFAD averaged over all CAPRICORN2 observations (Fig. 15a) shows a  
706 shadowy boomerang shape with a horizontal arm due to low clouds below 4 km and a  
707 diagonal arm due to deep convective clouds that extend beyond 6 km. The CAM6 COSP

708 CFAD (Fig. 15b) displays a shape analogous to observations but with much higher  
709 occurrence of reflectivities exceeding -10 dBZ. The upper arm of the AM4 COSP  
710 reflectivity CFAD is strongly shifted by ~25 dBZ toward reflectivities lower than  
711 observed (Fig. 15c).

712 Fig. 15 also shows separate CFADs for low vs. deep cloud columns, which are  
713 defined as having a maximum reflectivity above 4 km less (vs. greater) than -40 dBZ.  
714 The observed low-cloud CFAD (Fig. 15d) has a mode between -10 and 0 dBZ between 0-  
715 1 km in altitude associated with lightly precipitating cloud, with a lower tail extending to  
716 -40 dBZ contributed by low-level non-precipitating clouds. The CAM6 low-cloud CFAD  
717 (Fig. 15e) shows a comparable histogram of reflectivities, but with the maximum  
718 occurrence frequency at a slightly lower reflectivity near -10 dBZ and no tail of  
719 reflectivities below -20 dBZ and 1 km altitude. The AM4 low-cloud CFAD (Fig. 15f) is  
720 fairly similar to observations below 1 km altitude but underestimates reflectivities above  
721 1 km altitude.

722 The observed deep-cloud CFAD (Fig. 15g) constitutes the broader upper arm of  
723 the boomerang, with typical reflectivities clustering around 0 dBZ below 4 km and  
724 decreasing to ~ -20 dBZ at ~6 km (Fig. 15g). The CAM6 deep clouds (Fig. 15h) cluster at  
725 a comparable reflectivity range but occur more frequently than observed. Larger  
726 reflectivities are maintained at a much higher altitude in CAM6 as well. The AM4 deep  
727 clouds (Fig. 15i) have a -15 dBZ low bias in reflectivity except near the surface, where  
728 they are comparable in frequency and magnitude to observations.

729

#### 730 **4.4 Hydrometer microphysics inferred from COSP reflectivity decomposition**

731 *4.4.1 CAM6*

732 To better understand the contributions of different hydrometeors in CAM6 to  
733 reflectivity, we partition the non-attenuated COSP synthetic reflectivity into contributions  
734 from cloud liquid, cloud ice, rain and snow. Here we only consider large-scale  
735 precipitation, since convective precipitation rarely occurs in CAM6 along the ship track.  
736 The synthetic reflectivities of liquid, ice and rain are calculated from their respective grid  
737 mean number concentrations and effective radii following the formulas in COSP. The  
738 synthetic snow reflectivity is computed as the residual of the total nonattenuated COSP  
739 reflectivity and the sum of synthetic reflectivities from the other three hydrometers. AM4  
740 only outputs an attenuated reflectivity which cannot be exactly partitioned in this way.

741 We decompose the CAPRICORN2 CAM6 CFADs into cloud liquid, cloud ice,  
742 rain, and snow for all clouds (Fig. 16 a-d), low clouds (Fig. 16 e-h), and deep clouds (Fig.  
743 16 i-l). In all cases, stronger reflectivities are dominated by snow. CAM6 also simulates  
744 a substantial amount of cloud liquid with reflectivity below -20 dBZ and drizzle with  
745 reflectivity below -20 dBZ at altitudes below 2 km (Fig. 16a, e, i). Above 2 km, cloud ice  
746 becomes more prevalent in CAM6 but has low reflectivity below -10 dBZ. However,  
747 such low reflectivity is missing in the non-partitioned reflectivity (Figs. 15b, e, and h)  
748 suggesting that snow is more frequent in CAM6 than in the observations. The missing tail  
749 of low reflectivities might be also partly due to the insufficient sub-grid variability of  
750 cloud and precipitation in CAM6 COSP such that almost all simulated clouds have  
751 precipitation dominating their reflectivity.

752 The snow mass or size in CAM6 low clouds appears underestimated since its  
753 maximum frequency (Fig. 16h) is located at a lower reflectivity than the observations

754 (Fig. 15d). This indicates that snow in CAM6 low clouds is more homogeneous but less  
755 intense compared to the observations. For deep clouds, the frequency of occurrence of  
756 snow (Fig. 16l) is much higher than observations, while the grid average reflectivity is  
757 similar to observed at  $\sim 0$ dBZ. This implies that the snow in CAM6 deep clouds is  
758 similarly homogeneous and moderate. Note that the high snow occurrence could partially  
759 be attributed to the insufficient sub-grid variability of cloud and precipitation in CAM6  
760 COSP as mentioned earlier.

761

#### 762 4.4.2 AM4

763 To better understand the representation of hydrometeors in AM4, we compare  
764 time-height sections of grid mean liquid water and ice mixing ratios and precipitation  
765 fluxes from CAM6 and AM4 (Fig. 17). Normally AM4 shows substantially more cloud  
766 ice compared to CAM6 (Fig. 17f compared to b). The reason is that its microphysics  
767 scheme does not distinguish snow from ice and the cloud ice in AM4 is the sum of ice  
768 and snow. The AM4 downward ice flux is vertically continuous with the rain flux (Fig.  
769 17 g to c), confirming that above the freezing level the AM4 precipitation from deep and  
770 shallow clouds is in the form of sedimenting cloud ice particles. The snow flux  
771 approximated from the clear-sky ice flux as used in AM4 COSP (Fig. 17h) is less  
772 frequent and intense compared to the snow flux in CAM6 (Fig. 17d). AM4 has less  
773 supercooled liquid water above 2 km than CAM6 (Fig. 17 e to a), but our CAPRICORN2  
774 and SOCRATES observational analyses cannot as yet clearly test which model is closer  
775 to the truth.

776 To evaluate the snow intensity in AM4, we compare the hydrometeor PSDs in  
777 AM4 COSP with CAM6 COSP (Fig. 18). Here the PSDs are computed from area-  
778 weighted mean cloud liquid, cloud ice, rain and snow. AM4 has greater ice with much  
779 less rain and snow. Compared with CAM6 COSP snow PSDs, AM4 COSP significantly  
780 underestimates large snow particles with radius greater than 100 microns, leading to  
781 lower reflectivities. The AM4 COSP snow PSD is not taken from the AM4 microphysics,  
782 which would give no separate snow contribution to the PSD and worsen the AM4  
783 underestimate of reflectivity.

784

## 785 **5. Summary**

786 Observations of cloud properties from sophisticated in-situ and ship-based remote and  
787 in-situ sensors over the Southern Ocean during airborne (SOCRATES) and ship-based  
788 (CAPRICORN2) measurement campaigns during Jan.-Feb. 2018 are used to evaluate two  
789 state of the art atmospheric general circulation models (GCMs): CAM6 and AM4. These  
790 GCMs were nudged to reanalysis wind and temperature fields to minimize differences  
791 between modeled and observed synoptic conditions.

792 These measurements, together with collocated CERES TOA radiative flux estimates,  
793 provide a valuable dataset for evaluating simulations of cloud and precipitation in CAM6  
794 and AM4 and to understand their radiation biases. The major conclusions and implications  
795 are:

- 796 1. The nudged-meteorology simulation method facilitates detailed comparison of  
797 measured and simulated cloud properties from a limited set of observations in a  
798 synoptically variable environment.

- 799 2. Both GCMs correctly simulate that Southern Ocean supercooled boundary-layer  
800 clouds in that they reproduce observed compositions (i.e. they are mostly  
801 composed of small cloud droplets and larger precipitating ice particles).
- 802 3. CAM6 has too much cloud and that cloud is too bright (“too frequent, too  
803 bright”).
- 804 4. Cloud droplet number concentration in CAM6 is typically too low.
- 805 5. Precipitation in CAM6 is too frequent and too homogeneous.
- 806 6. AM4 has too little cloud occurrence, but the clouds are too bright (“too few, too  
807 bright”).
- 808 7. AM4 clouds include too much small ice and too little snow.

809

810 The low bias in cloud droplet number concentration in CAM6 is consistent with  
811 discrepancies seen between other state of the art models and satellite observations of  
812 Southern Ocean cloud droplet number concentrations in summertime low clouds (McCoy  
813 et al. 2020a in review, Revell et al. 2019). This low bias is a widespread issue remaining  
814 in GCMs that presumably contributes to TOA SW bias for low-lying liquid clouds over  
815 the Southern Ocean.

816 Both CAM6 COSP and AM4 COSP make assumptions about microphysics, size  
817 distributions, and horizontal homogeneity that are not fully consistent with their host  
818 GCM. Ideally such assumptions should be minimized, but at a minimum they must be  
819 kept in mind when comparing cloud radar data with COSP output. CAM6 COSP seems  
820 to simulate too large an area fraction of snow. AM4 simulates snow as a tail of the cloud  
821 ice distribution, while COSP expects a separate snow category. With or without COSP,

822 this results in AM4 simulating snow crystals that are too small and have far too little  
823 radar reflectivity.

824 The biggest challenge is still ahead – how to use the insights from this  
825 comprehensive analysis to improve the participating GCMs and their COSP simulators.  
826 We hope that the approach presented here will prove beneficial in testing other GCMs  
827 and developing improvements for future GCM versions.

828

829

830

### 831 **Appendix A: HSRL backscatter coefficient threshold in determining cloud** 832 **occurrence**

833 HSRL obtains the lidar return signal with high spectral resolution (<75 MHz laser  
834 bandwidth), which enables the separation of aerosol and cloud returns from molecular  
835 returns. Here we further separate cloud from aerosol returns by use of calibrated HSRL  
836 aerosol and cloud backscatter coefficient.

837

838 Examining the probability density function of HSRL cloud and aerosol  
839 backscatter coefficient for all 15 flights during SOCRATES (Fig. A1), we find a tri-  
840 modal distribution with three peaks locating near  $10^{-7}$ ,  $10^{-6}$ , and  $10^{-3} \text{ m}^{-1}\text{sr}^{-1}$   
841 respectively. Through inspection of HSRL lidar backscatter profiles (e.g., Figs. 4a  
842 and 4b), we interpret the two left modes as being contributed by the aerosols within  
843 and outside of the boundary layer, which are associated with lower backscatter  
844 coefficient than the rightmost cloud mode. We determine  $3 \times 10^{-5} \text{ m}^{-1}\text{sr}^{-1}$  as a HSRL

845 backscatter coefficient threshold separating the cloud mode from the two aerosol  
846 modes (the blue line in Fig. A1). This threshold was determined by a sensitivity test  
847 where we compare the HCR and HSRL cloud detection using different HSRL  
848 backscatter thresholds ranging from  $10^{-5}$  to  $10^{-4}$   $\text{m}^{-1}\text{sr}^{-1}$ . We find that the frequency of  
849 cloud occurrence as detected by HSRL is not sensitive to the threshold, but reduces  
850 quickly once the threshold increases beyond  $3 \times 10^{-5}$   $\text{m}^{-1}\text{sr}^{-1}$ .

851

## 852 **Appendix B: Droplet size distribution in CAM6 microphysics scheme and CAM6**

### 853 **COSP**

854

855 Use of CFADs as an observational constraint on GCM snowfall rate is complicated  
856 because the hydrometeor size distributions assumed in COSP do not match the internal  
857 distributions within the GCM microphysics. Here we compare CAM6 and CAM6 COSP  
858 DSDs for low clouds during CAPRICORN2 based on their respective hydrometeor size  
859 distribution assumptions described in Section 2.7 (Fig. 18). The hydrometeor PSDs are  
860 computed from their fraction mean masses and effective radii. Here we compare CAM6  
861 microphysics and CAM6 COSP here, since AM4 COSP snow is not taken from the AM4  
862 microphysics.

863

864 Rain and snow DSDs are represented well in CAM6 COSP. COSP slightly  
865 underestimates cloud liquid and overestimates ice particles, which leads to an  
866 underestimation (overestimation) in liquid (ice) reflectivities. However, this bias is not  
867 expected to significantly alter the net synthetic reflectivities in the frequently

868 precipitating CAM6 mixed–phased low clouds during CAPRICORN2 where snow  
869 dominates the reflectivity. A discrepancy is found for snow DSDs between CAM6 and  
870 CAM6 COSP, where CAM6 COSP has a greater concentration of small snowflakes (Fig.  
871 18d). We note that this discrepancy is caused by the inconsistency in snow densities  
872 assumed in CAM6 and CAM6 COSP. CAM6 COSP assumes a snow density of 100  
873  $\text{kg/m}^3$ , but the effective radius used by CAM6 COSP is computed in CAM6 by assuming  
874 a snow density of 250  $\text{kg/m}^3$ . The bigger snow density leads to a smaller effective radius,  
875 and therefore more small snowflakes and less big ones. Such discrepancy vanishes when  
876 the snow effective radius input into COSP is computed using a snow density of 100  
877  $\text{kg/m}^3$  (not shown). The density inconsistency barely affects the large particle number and  
878 has little impact on reflectivity.

879

880 It is reasonable to assume that the snow size distributions during CAPRICORN2 are  
881 similar to that during SOCRATES. Comparing Figs. 8 and 18 suggests that the mean  
882 snow PSD in CAM6 including all cloud types in SOCRATES is on average qualitatively  
883 consistent with the mean SOCRATES-observed DSD for precipitating low clouds,  
884 although the frequency of occurrence of snow is much higher.

885

886 **Acknowledgement:**

887 This work is funded by U.S. National Science Foundation (NSF) award numbers AGS-  
888 1660604 and AGS-1660609. The authors thank the National Center for Atmospheric  
889 Research (NCAR, supported by NSF) Earth Observing Laboratory (EOL) and CSIRO  
890 Marine National Facility (MNF) for supporting and undertaking the SOCRATES and

891 CAPRICORN2 deployment. We acknowledge the teams of SOCRATES and  
892 CAPRICORN2 scientists and technicians who made this work possible by collecting the  
893 data and maintaining the instruments. SOCRATES data are provided by EOL at  
894 <https://data.eol.ucar.edu/>. We thank Alain Protat (alain.protat@bom.gov.au) for providing  
895 radar reflectivity data during CAPRICORN2. CIRES SYN data used in this study were  
896 obtained from the NASA Earth Science Data Systems program:  
897 <https://search.earthdata.nasa.gov>.

898

## 899 **References**

900 Abdelmotaleb, A., Järvinen, E., Duft, D., Hirst, E., Vogt, S., Leisner, T., & Schnaiter, M. (2016). PHIPS-  
901 HALO: the airborne Particle Habit Imaging and Polar Scattering probe-Part 1: Design and operation.  
902 *Atmospheric Measurement Techniques*.

903

904 Atlas R. L., Bretherton C. S., and P. N. Blossey (2020) How well do high and low resolution models  
905 represent observed boundary layer structures and low clouds over the summertime Southern Ocean?  
906 *Journal of Geophysical Research: Atmospheres (submitted)*.

907

908 Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J. L., Klein, S. A., ... & John, V. O.  
909 (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American*  
910 *Meteorological Society*, **92**(8), 1023-1043.

911

912 Bodas-Salcedo, A., Williams, K. D., Ringer, M. A., Beau, I., Cole, J. N., Dufresne, J. L., ... & Yokohata, T.  
913 (2014). Origins of the solar radiation biases over the Southern Ocean in CFMIP2 models. *Journal of*  
914 *Climate*, **27**(1), 41-56.

915

916 Bodas-Salcedo, A., Mulcahy, J. P., Andrews, T., Williams, K. D., Ringer, M. A., Field, P. R., & Elsaesser,  
917 G. S. (2019). Strong dependence of atmospheric feedbacks on mixed-phase microphysics and aerosol-  
918 cloud interactions in HadGEM3. *Journal of Advances in Modeling Earth Systems*.  
919  
920  
921 Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., & Schanen, D. P. (2013).  
922 Higher-order turbulence closure and its impact on climate simulations in the Community Atmosphere  
923 Model. *Journal of Climate*, **26**(23), 9655-9676.  
924  
925 Bogenschutz, P. A., Gettelman, A., Hannay, C., Larson, V. E., Neale, R. B., Craig, C., & Chen, C. C.  
926 (2018). The path to CAM6: Coupled simulations with CAM5. 4 and CAM5. 5. *Geoscientific Model*  
927 *Development (Online)*, *11*(LLNL-JRNL-731418).  
928  
929 Bony, S., & Dufresne, J. L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback  
930 uncertainties in climate models. *Geophysical Research Letters*, *32*(20).  
931  
932 Bretherton, C. S., McCoy, I. L., Mohrmann, J., Wood, R., Ghate, V., Gettelman, A., ... & Zuidema, P.  
933 (2019). Cloud, aerosol, and boundary layer structure across the northeast Pacific stratocumulus-  
934 cumulus transition as observed during CSET. *Monthly Weather Review*, *147*(6), 2083-2103.  
935  
936 Ceppi, P., Hwang, Y. T., Frierson, D. M., & Hartmann, D. L. (2012). Southern Hemisphere jet latitude  
937 biases in CMIP5 models linked to shortwave cloud forcing. *Geophysical Research Letters*, *39*(19).  
938  
939 Cess, R. D., Zhang, M. H., Ingram, W. J., Potter, G. L., Alekseev, V., Barker, H. W., ... & Dix, M. R.  
940 (1996). Cloud feedback in atmospheric general circulation models: An update. *Journal of Geophysical*  
941 *Research: Atmospheres*, *101*(D8), 12791-12794.  
942

943 Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., ... & Sun, M. (2013).  
944 Geostationary enhanced temporal interpolation for CERES flux products. *Journal of Atmospheric and*  
945 *Oceanic Technology*, 30(6), 1072-1090.

946

947 Donner, L. J., Seman, C. J., Soden, B. J., Hemler, R. S., Warren, J. C., Ström, J., & Liou, K. N. (1997).  
948 Large-scale ice clouds in the GFDL SKYHI general circulation model. *Journal of Geophysical*  
949 *Research: Atmospheres*, 102(D18), 21745-21768.

950

951 Dunne et al. (2019). Model description and simulation characteristics. *Journal of Advances in Modeling*  
952 *Earth Systems*. (submitted)

953

954 UCAR/NCAR - Earth Observing Laboratory. (2005). NSF/NCAR GV HIAPER Aircraft. UCAR/NCAR -  
955 Earth Observing Laboratory. <https://doi.org/10.5065/D6DR2SJP> Retrieved December 14, 2016

956

957 SouthWest Sciences, I. (SWS), & UCAR/NCAR - Earth Observing Laboratory. (2008). Vertical Cavity  
958 Surface-Emitting Laser (VCSEL) Hygrometer. UCAR/NCAR - Earth Observing Laboratory.  
959 <https://doi.org/10.5065/D6PV6HDM> Retrieved December 20, 2016

960

961 UCAR/NCAR - Earth Observing Laboratory, & University of Wisconsin. (2010). High Spectral Resolution  
962 Lidar for the Gulfstream-V; G5-HSRL. UCAR/NCAR - Earth Observing Laboratory.  
963 <https://doi.org/10.5065/D67W6976> Retrieved December 20, 2016

964

965 UCAR/NCAR - Earth Observing Laboratory. (2014). HIAPER Cloud Radar (HCR). UCAR/NCAR - Earth  
966 Observing Laboratory. <https://doi.org/10.5065/D6BP00TP> Retrieved December 20, 2016

967

968 Engström, A., Bender, F. M., Charlson, R. J., & Wood, R. (2015). The nonlinear relationship between  
969 albedo and cloud fraction on near-global, monthly mean scale in observations and in the CMIP5 model  
970 ensemble. *Geophysical Research Letters*, 42(21), 9571-9578.

971  
972  
973 Eyring, V., Bony, S., Meehl, G. A., Senior, C., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2015).  
974 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and  
975 organisation. *Geoscientific Model Development Discussions*, 8(12).  
976  
977  
978 Gettelman, A., & Morrison, H. (2015). Advanced two-moment bulk microphysics for global models. Part I:  
979 Off-line tests and comparison with other schemes. *Journal of Climate*, 28(3), 1268-1287.  
980  
981 Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., ... &  
982 Mills, M. J. (2019). High climate sensitivity in the Community Earth System Model Version 2  
983 (CESM2). *Geophysical Research Letters*.  
984  
985 Gettelman, A., Bardeen C. G., McCluskey C. S., Jävinen E., Stith J., Bretherton C. (2020). Simulating  
986 Observations of Southern Ocean Clouds and Implications for Climate. *Journal of Geophysical*  
987 *Research: Atmospheres*. (submitted)  
988  
989 Gordon, N. D., & Klein, S. A. (2014). Low-cloud optical depth feedback in climate models. *Journal of*  
990 *Geophysical Research: Atmospheres*, 119(10), 6052-6065.  
991  
992 Guo, H., Golaz, J. C., Donner, L. J., Ginoux, P., & Hemler, R. S. (2014). Multivariate probability density  
993 functions with dynamics in the GFDL atmospheric general circulation model: Global tests. *Journal of*  
994 *Climate*, 27(5), 2087-2108.  
995  
996 Guo, H., Golaz, J. C., Donner, L. J., Wyman, B., Zhao, M., & Ginoux, P. (2015). CLUBB as a unified  
997 cloud parameterization: Opportunities and challenges. *Geophysical Research Letters*, 42(11), 4540-  
998 4547.

999

1000 Haynes, J. M., Marchand, R. T., Luo, Z., Bodas-Salcedo, A., & Stephens, G. L. (2007). A multipurpose  
1001 radar simulation package: QuickBeam. *Bulletin of the American Meteorological Society*, 88(11), 1723-  
1002 1728.

1003

1004 Held, I. M., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *J. Adv. Model.*  
1005 *Earth Syst.*, 11, 3691– 3727. <https://doi.org/10.1029/2019MS001829>

1006

1007 Hersbach, H., and D. Dee, 2016: ERA-5 reanalysis is in pro-duction.ECMWF Newsletter, No. 147,  
1008 ECMWF, Reading, United Kingdom,<https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis->  
1009 production

1010

1011 Hoose, C., Kristjansson, J. E., Chen, J.-P., & Hazra, A. (2010, March). A Classical-Theory-Based  
1012 Parameterization of Heterogeneous Ice Nucleation by Mineral Dust, Soot, and Biological Particles in a  
1013 Global Climate Model. *J. Atmos. Sci.*, 67 (8), 2483-2503. doi: 10.1175/2010JAS3425.1

1014

1015 Houze Jr, R. A., Wilton, D. C., & Smull, B. F. (2007). Monsoon convection in the Himalayan region as  
1016 seen by the TRMM Precipitation Radar. *Quarterly Journal of the Royal Meteorological Society: A*  
1017 *journal of the atmospheric sciences, applied meteorology and physical oceanography*, 133(627), 1389-  
1018 1411.

1019

1020 Hwang, Y. T., Frierson, D. M., & Kang, S. M. (2013). Anthropogenic sulfate aerosol and the southward  
1021 shift of tropical precipitation in the late 20th century. *Geophysical Research Letters*, 40(11), 2845-  
1022 2850.

1023

1024 Jeuken, A. B. M., Siegmund, P. C., Heijboer, L. C., Feichter, J., & Bengtsson, L. (1996). On the potential  
1025 of assimilating meteorological analyses in a global climate model for the purpose of model validation.  
1026 *Journal of Geophysical Research: Atmospheres*, 101(D12), 16939-16950.

1027

1028 Kay, J. E., Wall, C., Yettella, V., Medeiros, B., Hannay, C., Caldwell, P., & Bitz, C. (2016). Global climate  
1029 impacts of fixing the Southern Ocean shortwave radiation bias in the Community Earth System Model  
1030 (CESM). *Journal of Climate*, 29(12), 4617-4636.

1031

1032

1033 Kollias, P., Clothiaux, E. E., Miller, M. A., Albrecht, B. A., Stephens, G. L., & Ackerman, T. P. (2007).  
1034 Millimeter-wavelength radars: New frontier in atmospheric cloud and precipitation research. *Bulletin*  
1035 *of the American Meteorological Society*, 88(10), 1608-1624.

1036

1037 Liu, X., Ma, P. L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., ... & Rasch, P. J. (2016). Description and  
1038 evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within version 5.3 of  
1039 the Community Atmosphere Model. *Geoscientific Model Development (Online)*, 9(PNNL-SA-  
1040 110649).

1041

1042 Mace, G. G., & Zhang, Q. (2014). The CloudSat radar-lidar geometrical profile product (RL-GeoProf):  
1043 Updates, improvements, and selected results. *Journal of Geophysical Research: Atmospheres*, 119(15),  
1044 9441-9462.

1045

1046 Mace, G. G., & Protat, A. (2018). Clouds over the Southern Ocean as observed from the R/V Investigator  
1047 during CAPRICORN. Part I: Cloud occurrence and phase partitioning. *Journal of Applied Meteorology*  
1048 *and Climatology*, 57(8), 1783-1803.

1049

1050

1051 Mason, S., Jakob, C., Protat, A., & Delanoë, J. (2014). Characterizing observed midtopped cloud regimes  
1052 associated with Southern Ocean shortwave radiation biases. *Journal of Climate*, 27(16), 6189-6203.

1053

1054 Matrosov, S. Y. (2007). Potential for attenuation-based estimations of rainfall rate from CloudSat.  
1055 *Geophysical research letters*, 34(5).  
1056

1057 McCoy, D. T., Tan, I., Hartmann, D. L., Zelinka, M. D., & Storelvmo, T. (2016). On the relationships  
1058 among cloud cover, mixed-phase partitioning, and planetary albedo in GCMs. *Journal of Advances in*  
1059 *Modeling Earth Systems*, 8(2), 650-668.  
1060

1061 McCoy, I. L., McCoy D. T., Wood R. Regayre L., Watson-Parris D., Grosvenor D. P., Mulcahy J. P., Hu  
1062 Y., Bender F. A. –M., Field P. R., Carslaw K., Gordon H. (2020a). The hemispheric contrast in cloud  
1063 microphysical properties constrains aerosol forcing. *Proceedings of the National Academy of Sciences*  
1064 *of the United States of America*. Submitted.  
1065

1066 McCoy I. L., Bretherton C. S., Wood R., Twohy C. H., Gettelman A., Bardeen C. (2020b). Recent Particle  
1067 Formation and Aerosol Variability Near Southern Ocean Low Clouds. In preparation.  
1068

1069 Meehl, G. A., Covey, C., McAvaney, B., Latif, M., & Stouffer, R. J. (2005). Overview of the coupled  
1070 model intercomparison project. *Bulletin of the American Meteorological Society*, 86(1), 89-93.  
1071

1072 Ming, Y., Ramaswamy, V., Donner, L. J., & Phillips, V. T. J. (2005). A robust parameterization of cloud  
1073 droplet activation. *J. Atmos. Sci.*  
1074

1075 Ming, Y., Ramaswamy, V., Donner, L. J., Phillips, V. T., Klein, S. A., Ginoux, P. A., & Horowitz, L. W.  
1076 (2007). Modeling the interactions between aerosols and liquid water clouds with a self-consistent cloud  
1077 scheme in a general circulation model. *Journal of the Atmospheric Sciences*, 64(4), 1189-1209.  
1078

1079 Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric  
1080 general circulation model: Evolution from MERRA to MERRA2. *Geoscientific Model Development*,  
1081 8(5), 1339-1356.

1082  
1083 Nam, C., Bony, S., Dufresne, J. L., & Chepfer, H. (2012). The ‘too few, too bright’ tropical low-cloud  
1084 problem in CMIP5 models. *Geophysical Research Letters*, 39(21).  
1085  
1086 Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C., & Frey, R. A. (2003).  
1087 The MODIS cloud products: Algorithms and examples from Terra. *IEEE Transactions on Geoscience*  
1088 *and Remote Sensing*, 41(2), 459-473.  
1089  
1090 Protat, A., Schulz, E., Rikus, L., Sun, Z., Xiao, Y., & Keywood, M. (2017). Shipborne observations of the  
1091 radiative effect of Southern Ocean clouds. *Journal of Geophysical Research: Atmospheres*, 122(1),  
1092 318-328.  
1093  
1094  
1095 Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., ... & Bloom, S. (2011).  
1096 MERRA: NASA’s modern-era retrospective analysis for research and applications. *Journal of climate*,  
1097 24(14), 3624-3648.  
1098  
1099 Rotstayn, L. D. (1997). A physically based scheme for the treatment of stratiform clouds and precipitation  
1100 in large-scale models. I: Description and evaluation of the microphysical processes. *Quarterly Journal*  
1101 *of the Royal Meteorological Society*, 123(541), 1227-1282.  
1102  
1103  
1104 Rotstayn, L. D. (2000). On the “tuning” of autoconversion parameterizations in climate models. *Journal of*  
1105 *Geophysical Research: Atmospheres*, 105(D12), 15495-15507.  
1106  
1107  
1108

1109 Rutan, D. A., Kato, S., Doelling, D. R., Rose, F. G., Nguyen, L. T., Caldwell, T. E., & Loeb, N. G. (2015).  
1110 CERES synoptic product: Methodology and validation of surface radiant flux. *Journal of Atmospheric*  
1111 *and Oceanic Technology*, 32(6), 1121-1143.  
1112  
1113 Schnaiter, M. 2018. PHIPS-HALO Stereo Imaging Data. Version 1.0. UCAR/NCAR - Earth Observing  
1114 Laboratory. <https://doi.org/10.5065/D62B8WWF>. Accessed 14 Jan 2020.  
1115  
1116 Schnaiter, M., Järvinen, E., Ahmed, A., & Leisner, T. (2018). PHIPS-HALO: the airborne particle habit  
1117 imaging and polar scattering probe–Part 2: Characterization and first results. *Atmospheric*  
1118 *Measurement Techniques*, 11(1), 341.  
1119  
1120 Shi, X., Liu, X., & Zhang, K. (2015, February). Effects of pre-existing ice crystals on cirrus clouds and  
1121 comparison between different ice nucleation parameterizations with the Community Atmosphere  
1122 Model (CAM5). *Atmospheric Chemistry and Physics*, 15 (3), 1503-1520. doi: 10.5194/acp-15-1503-  
1123 2015  
1124  
1125 Swales, D. J., Pincus, R., & Bodas-Salcedo, A. (2018). The cloud feedback model intercomparison project  
1126 observational simulator package: Version 2. *Geoscientific Model Development*, 11(1), 77-81.  
1127  
1128 Song, H., Zhang, Z., Ma, P. L., Ghan, S., & Wang, M. (2018). The importance of considering sub-grid  
1129 cloud variability when using satellite observations to evaluate the cloud and precipitation simulations  
1130 in climate models. *UMBC Physics Department*.  
1131  
1132 Tan, I., Storelvmo, T., & Zelinka, M. D. (2016). Observational constraints on mixed-phase clouds imply  
1133 higher climate sensitivity. *Science*, 352(6282), 224-227.  
1134

1135 Terai, C. R., Klein, S. A., & Zelinka, M. D. (2016). Constraining the low-cloud optical depth feedback at  
1136 middle and high latitudes using satellite observations. *Journal of Geophysical Research: Atmospheres*,  
1137 *121*(16), 9696-9716.  
1138  
1139  
1140 Tiedtke, M. (1993). Representation of clouds in large-scale models. *Monthly Weather Review*, *121*(11),  
1141 3040-3061.  
1142  
1143 Trenberth, K. E., & Fasullo, J. T. (2010). Simulation of present-day and twenty-first-century energy  
1144 budgets of the southern oceans. *Journal of Climate*, *23*(2), 440-454.  
1145  
1146  
1147 UCAR/NCAR - Earth Observing Laboratory. (2005). NSF/NCAR GV HIAPER Aircraft. UCAR/NCAR -  
1148 Earth Observing Laboratory. <https://doi.org/10.5065/D6DR2SJP> Retrieved December 14, 2016.  
1149  
1150 Wang, Y., Liu, X., Hoose, C., & Wang, B. (2014, October). Different contact angle distributions for  
1151 heterogeneous ice nucleation in the Community Atmospheric Model version 5. *Atmos. Chem. Phys.*, *14*  
1152 (19), 10411-10430. doi:10.5194/acp-14-10411-2014  
1153  
1154 Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., ... &  
1155 Williamson, D. L. (2013). The Transpose-AMIP II experiment and its application to the understanding  
1156 of Southern Ocean cloud biases in climate models. *Journal of Climate*, *26*(10), 3258-3274.  
1157  
1158 Wexler, A. (1976). Vapor pressure formulation for water in range 0 to 100 C. A revision. *J. Res. Natl. Bur.*  
1159 *Stand. A*, *80*, 775-785.  
1160

1161 Wu, W., McFarquhar, G. 2019. NSF/NCAR GV HIAPER 2D-S Particle Size Distribution (PSD) Product  
1162 Data. Version 1.1. UCAR/NCAR - Earth Observing Laboratory. <https://doi.org/10.26023/8HMG->  
1163 [WQP3-XA0X](https://doi.org/10.26023/8HMG-WQP3-XA0X). Accessed 14 Jan 2020.

1164

1165 Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee III, R. B., Smith, G. L., & Cooper, J. E. (1996).  
1166 Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment.  
1167 *Bulletin of the American Meteorological Society*, 77(5), 853-868.

1168

1169 Wu, C., Liu, X., Diao, M., Zhang, K., Gettelman, A., Lu, Z., Penner, J. E., and Lin, Z., 2017: Direct  
1170 comparisons of ice cloud macro- and microphysical properties simulated by the Community  
1171 Atmosphere Model version 5 with HIPPO aircraft observations, *Atmos. Chem. Phys.*, **17**, 4731-4749,  
1172 <https://doi.org/10.5194/acp-17-4731-2017>.

1173

1174 Yuter, S. E., & Houze Jr, R. A. (1995). Three-dimensional kinematic and microphysical evolution of  
1175 Florida cumulonimbus. Part II: Frequency distributions of vertical velocity, reflectivity, and  
1176 differential reflectivity. *Monthly weather review*, 123(7), 1941-1963.

1177

1178 Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., ... & Taylor, K. E.  
1179 (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, 47(1),  
1180 e2019GL085782.

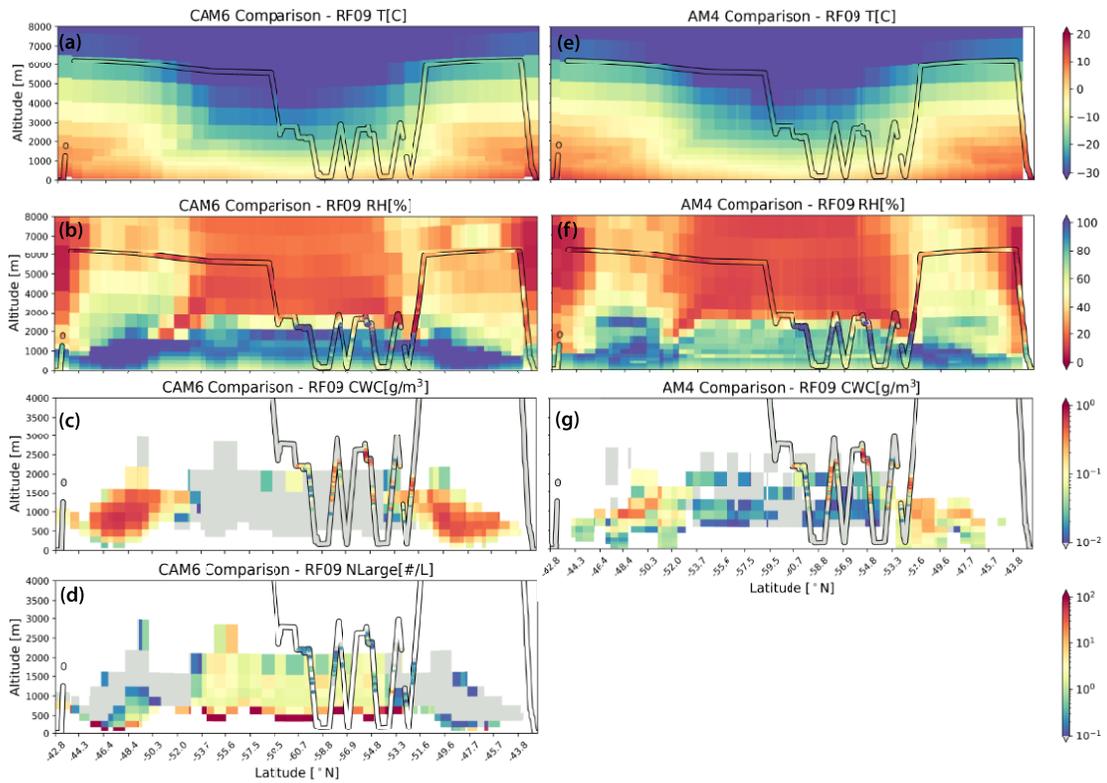
1181

1182 Zhao, M., Golaz, J. C., Held, I. M., Ramaswamy, V., Lin, S. J., Ming, Y., ... & Guo, H. (2016). Uncertainty  
1183 in model climate sensitivity traced to representations of cumulus precipitation microphysics. *Journal of*  
1184 *Climate*, 29(2), 543-560.

1185

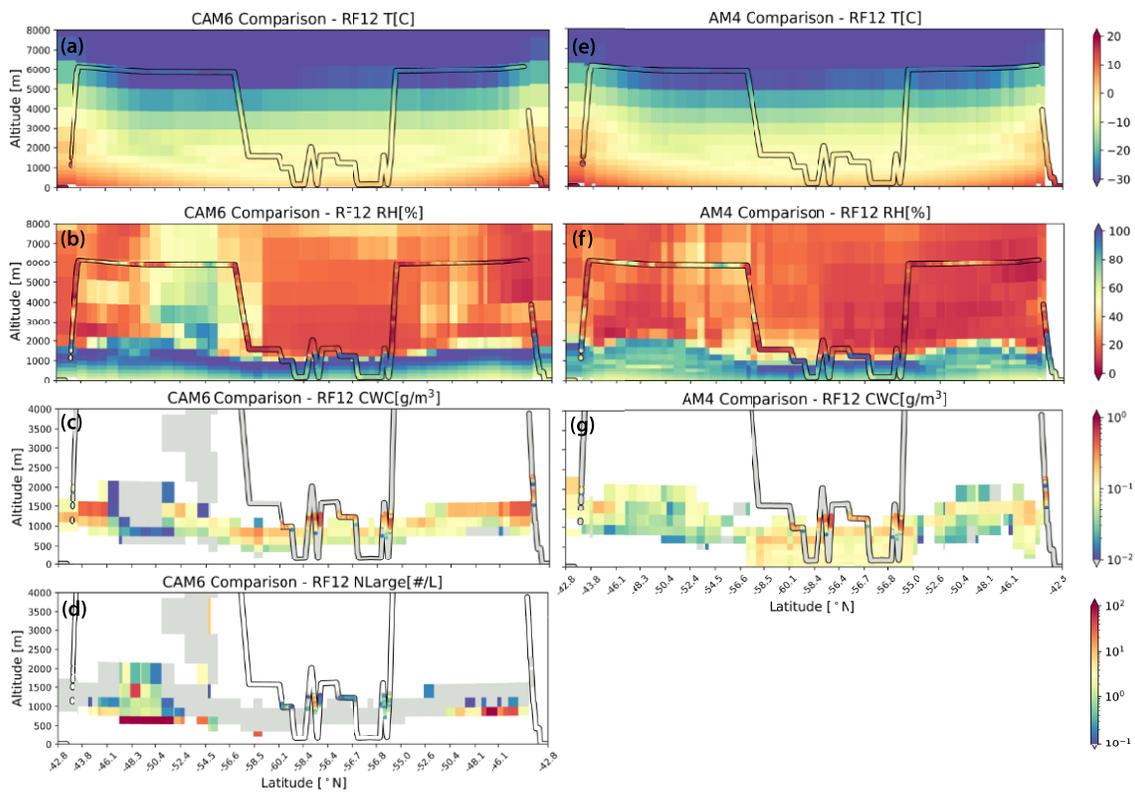
1186 Zhao, M., Golaz, J. C., Held, I. M., Guo, H., Balaji, V., Benson, R., ... & Dunne, K. (2018). The GFDL  
1187 global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning  
1188 strategies. *Journal of Advances in Modeling Earth Systems*, 10(3), 735-769.

1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201

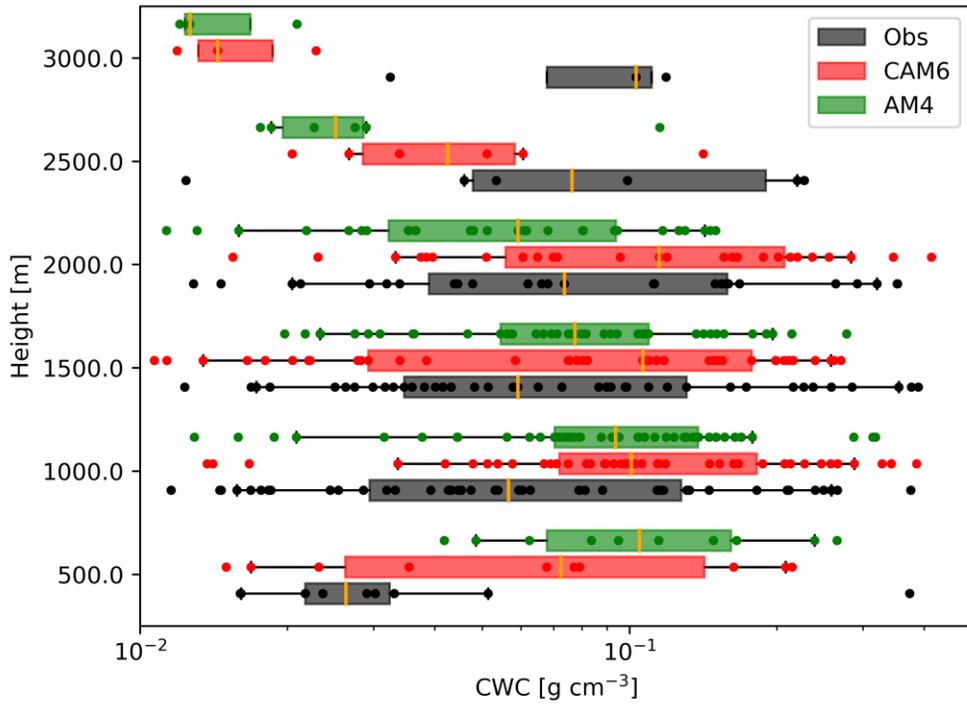


1202

1203 Fig. 1. SOCRATES flight RF09 observed (a) ambient temperature, (b) relative humidity,  
 1204 (c) liquid cloud water content from CDP, and (d) large particle number density N<sub>Large</sub> (a  
 1205 precipitation indicator described in the text), shown as shading within black channels,  
 1206 overlying the corresponding CAM6 model output. (e)-(g): same as (a)-(c) but overlying  
 1207 profiles in AM4. N<sub>Large</sub> cannot be computed from AM4 outputs. Gray shading denotes  
 1208 data that falls below the trusted value range. Missing data is shown either as gaps in the  
 1209 observation channel or in white.  
 1210  
 1211

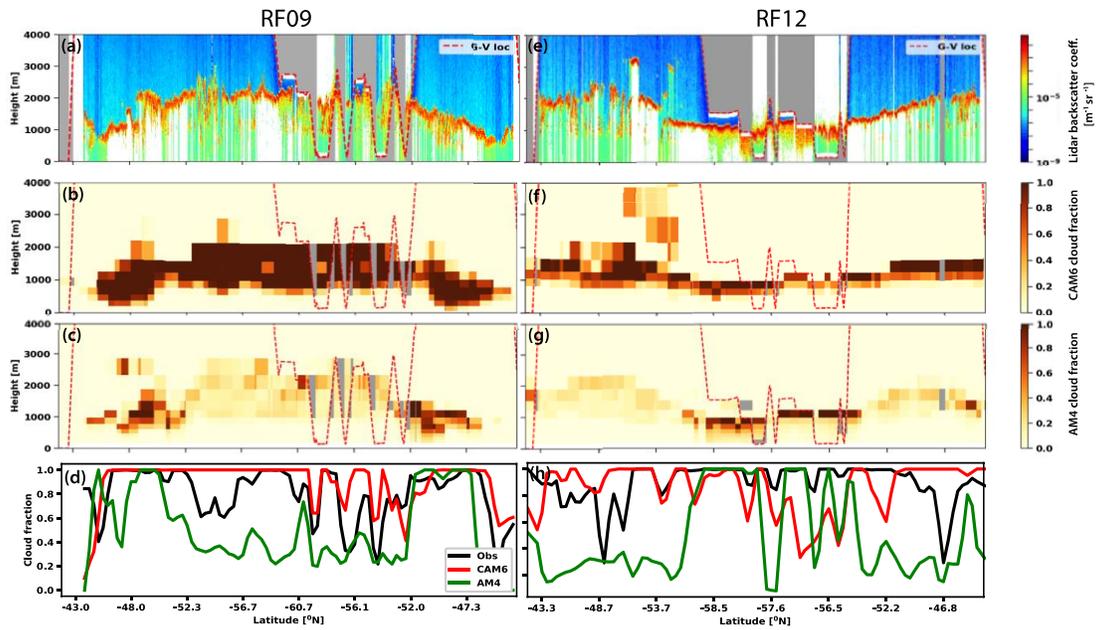


1212  
 1213 Fig. 2. Same as Fig. 1 but for flight RF12.



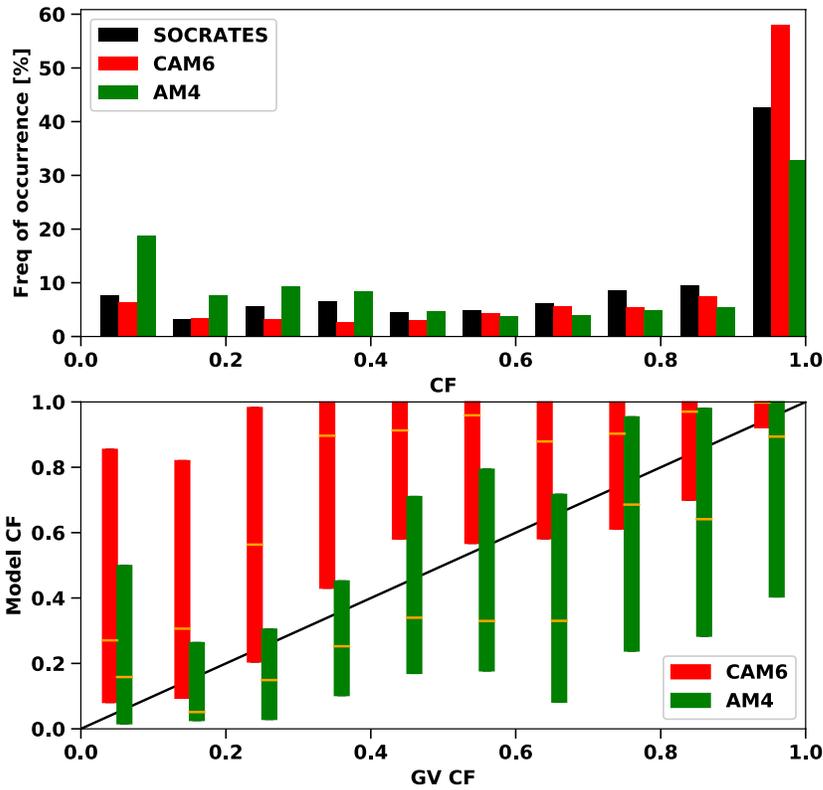
1214

1215 Fig. 3. Inter-quartile range boxes of observed and GCM in-cloud water content at  
 1216 different heights below 3 km for 15 flights during the SOCRATES campaign. Data is  
 1217 binned into boxes of 500 m in altitude and 25 minutes in time (dots) before range boxes  
 1218 are calculated. The orange bar inside the box indicates a median value for each bin and  
 1219 the whiskers indicate a range of 5-95%.



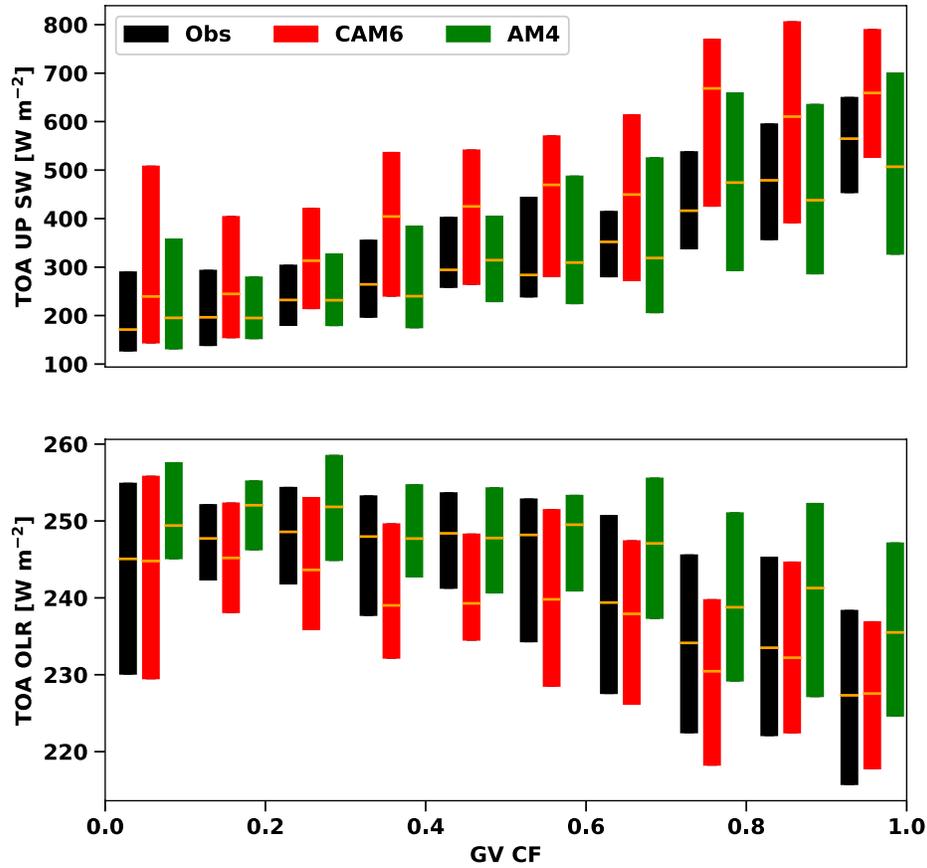
1220

1221 Fig. 4. (a) HSRL lidar backscatter coefficient, (b) CAM6 cloud fraction, (c) AM4 cloud  
 1222 fraction, and (d) cloud occurrence below 4 km from observations, in CAM6, and AM4  
 1223 for SOCRATES flight RF09. Red dashed lines indicate the position of the GV aircraft.  
 1224 Gray shading indicates the area of no observations. HSRL backscatter within the dead  
 1225 zone extending 150 m from the aircraft is masked white. (e)-(h), same as (a)-(d) but for  
 1226 RF12.



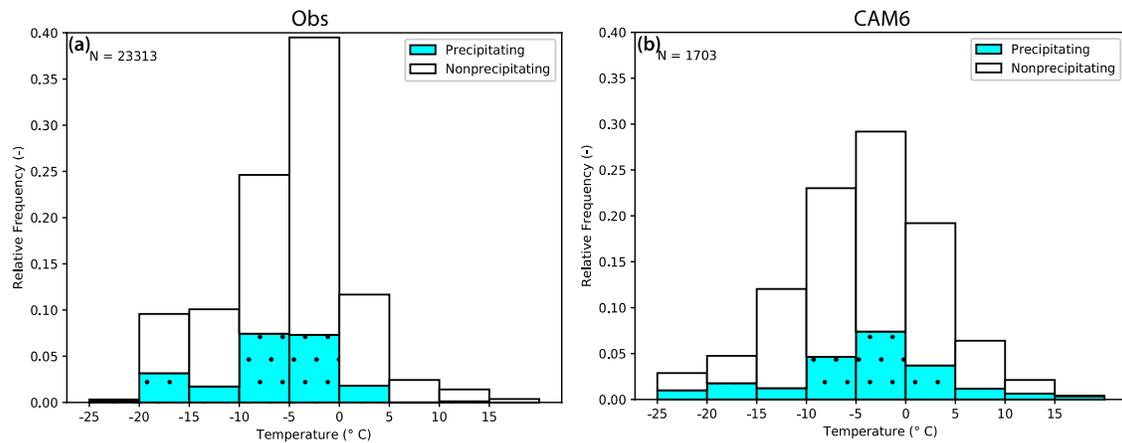
1227

1228 Fig. 5. (a) Frequency of occurrence of low clouds (below 4 km) from observations  
 1229 (black), CAM6 (red), and AM4 (green), and (b) inter-quartile range boxes of CAM6 and  
 1230 AM4 low cloud occurrence binned to 0.1 of observed low cloud fraction during  
 1231 SOCRATES. The orange bar inside each box indicates the bin-median.



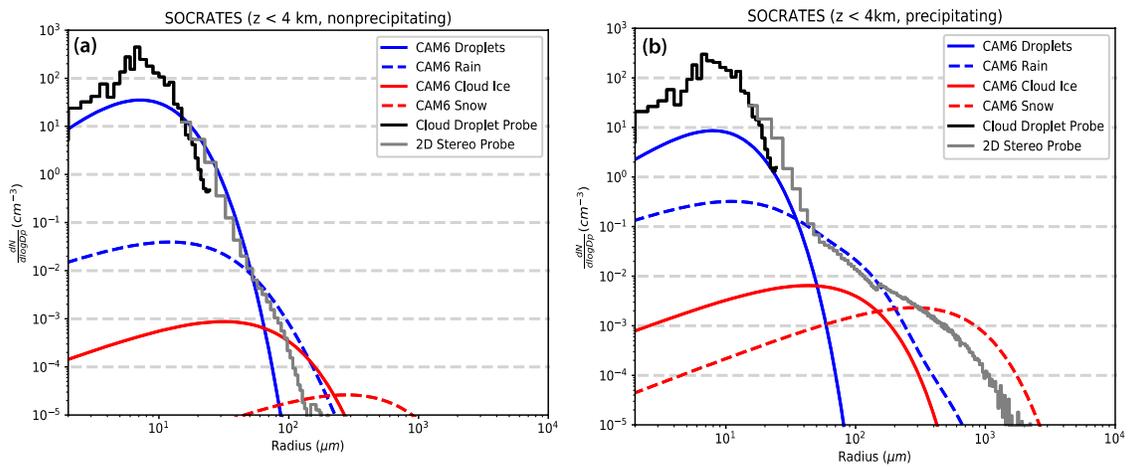
1232

1233 Fig. 6. Inter-quartile range boxes of observation matched CERES SYN (black), CAM6  
 1234 (red), and AM4 (green) TOA (a) RSW scaled to insolation at solar noon and (b) OLR,  
 1235 averaged over bins of observed low cloud occurrence during SOCRATES. The orange  
 1236 bar inside the box indicates a bin-median.

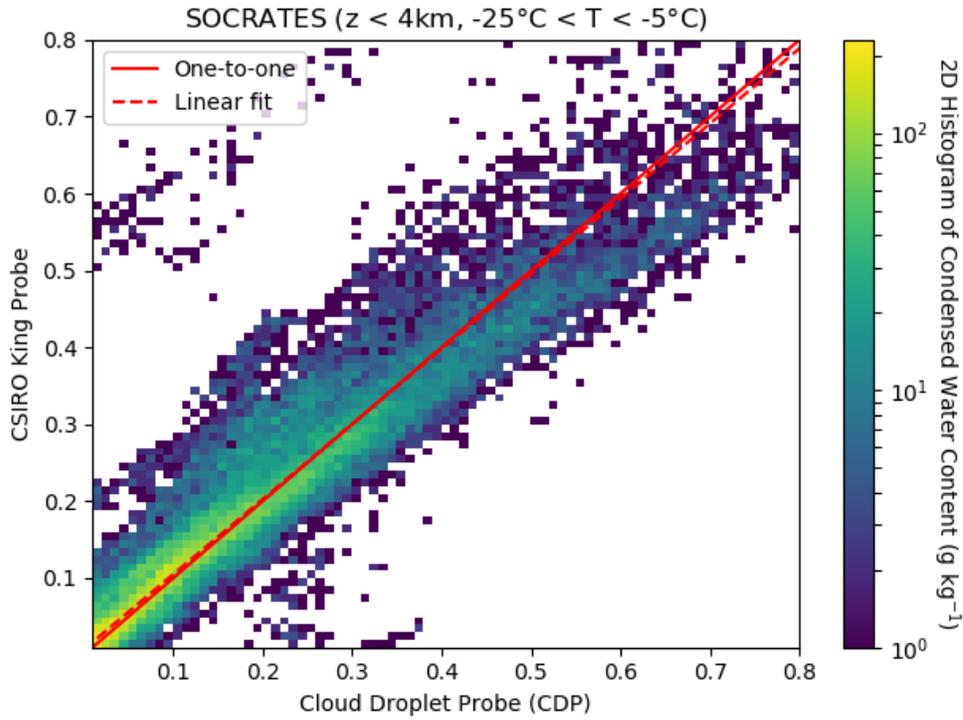


1237

1238 Fig. 7. Stacked histogram of occurrence frequency of (a) observed, and (b) CAM6  
 1239 nonprecipitating and precipitating low clouds along the SOCRATES flight tracks, sorted  
 1240 by ambient temperature.  
 1241



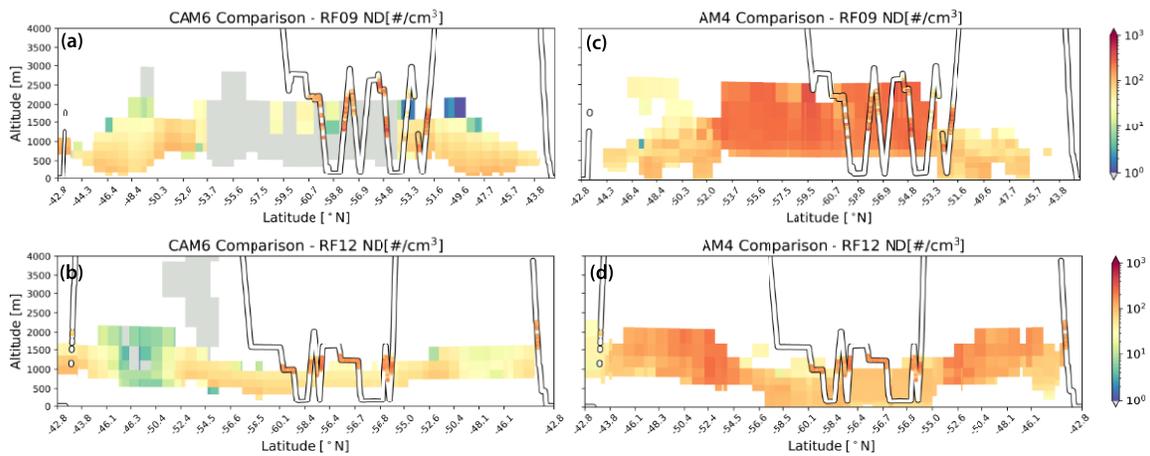
1242  
 1243 Fig. 8. Particle size distributions averaged across SOCRATES for CDP and 2DS  
 1244 observations (black lines) and CAM6 (colored lines) in (a) nonprecipitating low clouds,  
 1245 and (b) precipitating low clouds.



1246

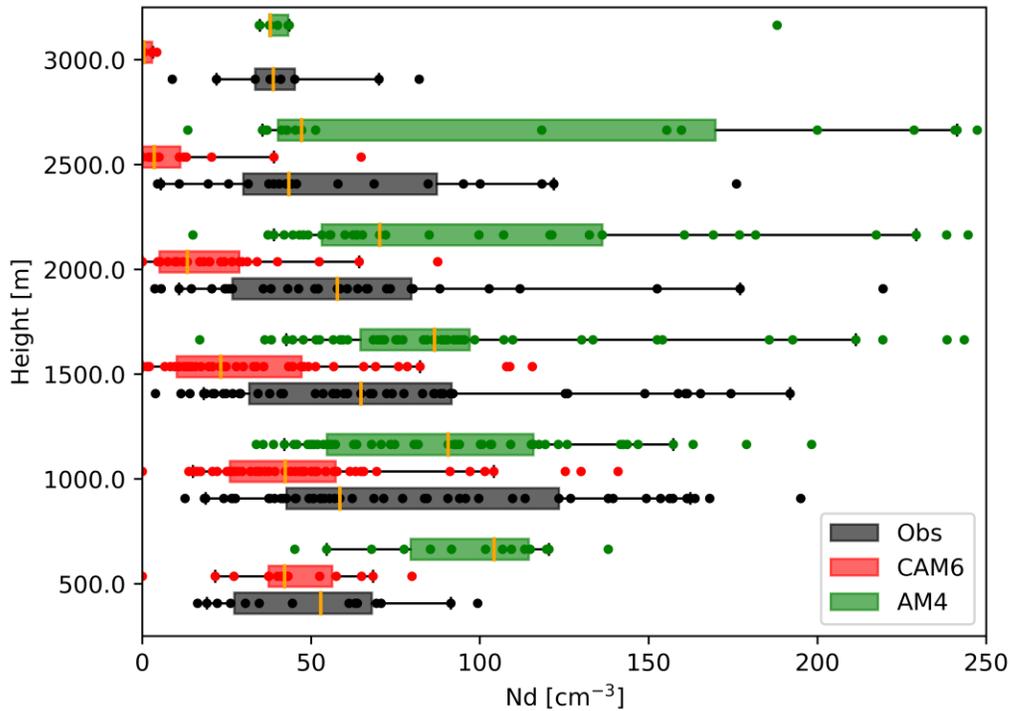
1247 Fig. 9 2D histogram of condensed water content for cloud droplet probe and CSIRO

1248 King Probe for low clouds at temperature between  $-5^{\circ}\text{C}$  and  $-25^{\circ}\text{C}$ .

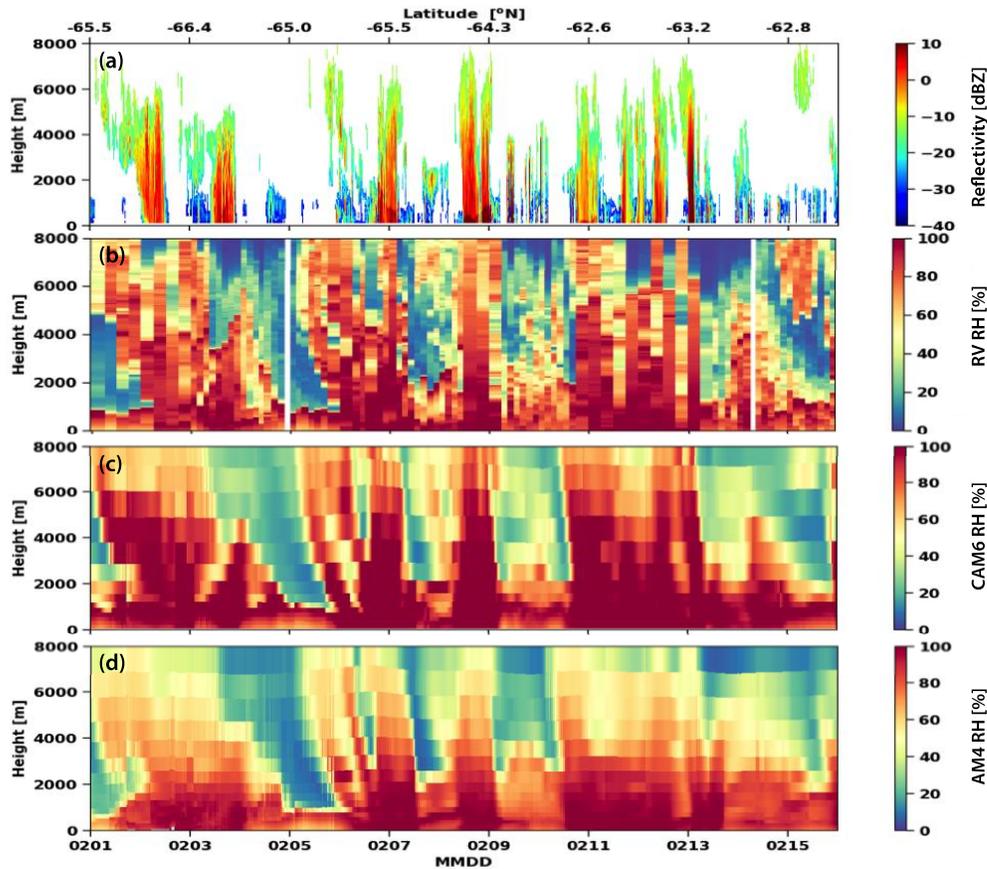


1249

1250 Fig. 10 In-cloud CDP-derived droplet number concentration for SOCRATES flight RF09  
 1251 and RF12 (shades inside black channels) overlying the corresponding variable profiles in  
 1252 CAM6 and AM4 as in Fig . 1 and 2.

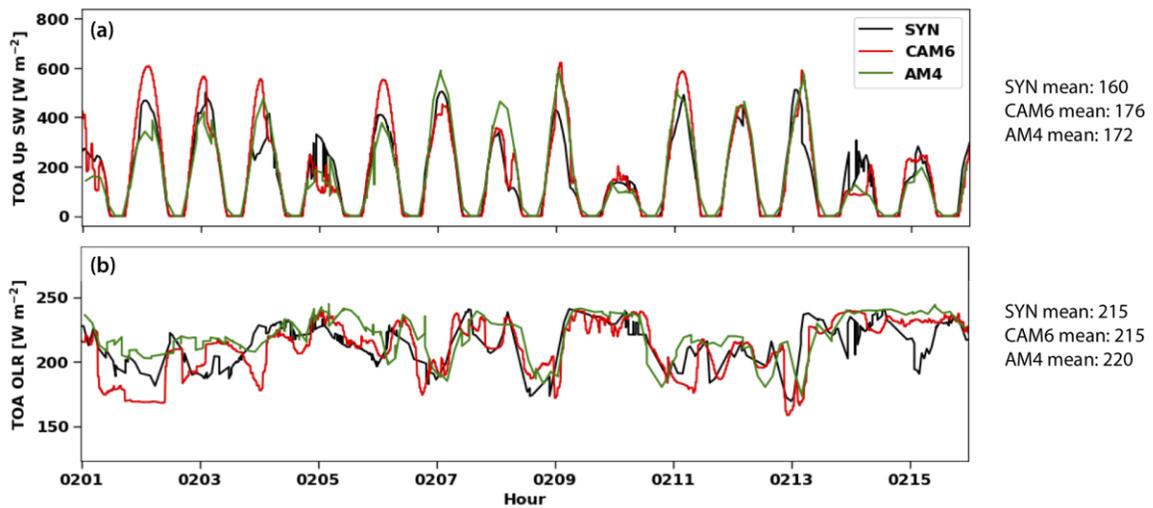


1253  
 1254 Fig. 11 Inter-quartile range boxes of observed and modeled in-cloud droplet number  
 1255 concentration binned into boxes of 500 m in altitude and 25 minutes in time (dots) before  
 1256 range boxes are calculated (as in Fig. 3). All in-cloud data ( $CWC > 0.01 \text{ g m}^{-3}$ ) up to 3 km  
 1257 across all 15 SOCRATES flights is included in the bin mean calculation. Only average  
 1258  $N_d \geq 1 \text{ cm}^{-3}$  is used in calculating the range boxes. The orange bar inside the box indicates  
 1259 a bin-median and the whiskers indicate the 5-95% range.



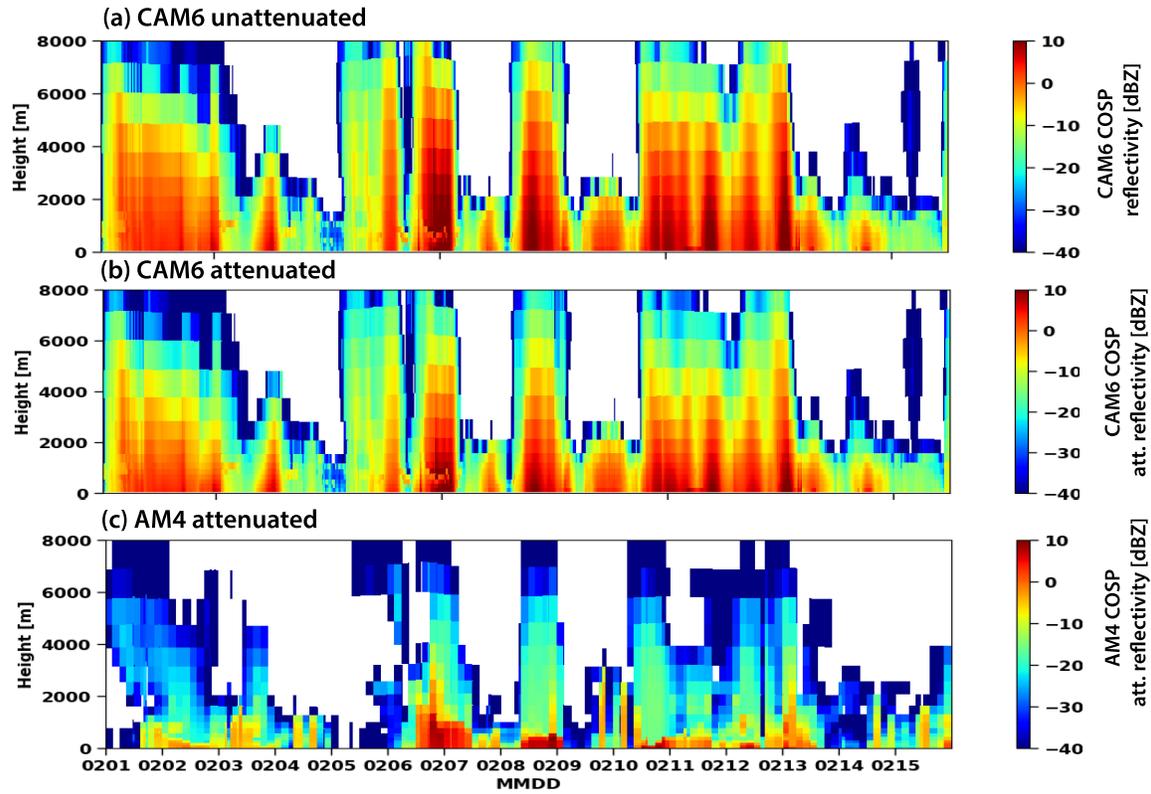
1260

1261 Fig. 12 (a) Ship-based upward-pointing W-band radar reflectivity, and relative humidity  
 1262 profiles from (b) radiosondes, (c) CAM6, and (d) AM4 during the 1-15 February 2018  
 1263 period of CAPRICORN2.

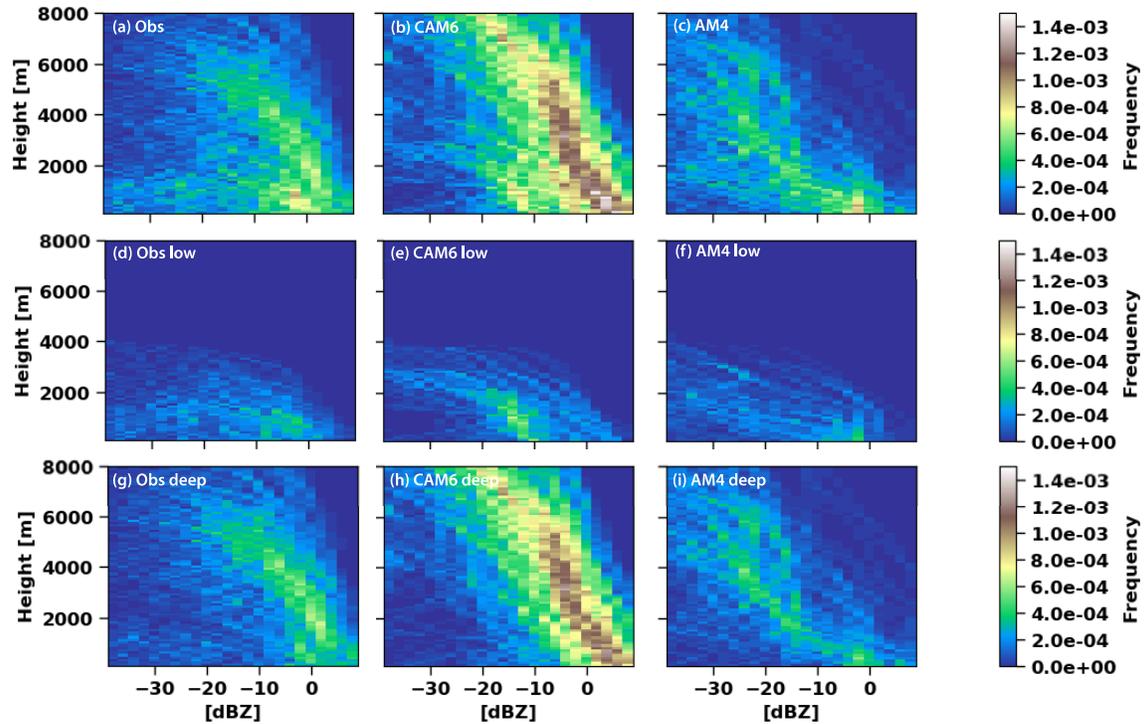


1264

1265 Fig. 13 TOA (a) RSW and (b) OLR from CERES SYN observations (black), CAM6  
1266 (red), and AM4 (green) during the 1-15 February 2018 period of the CAPRICORN2  
1267 campaign.



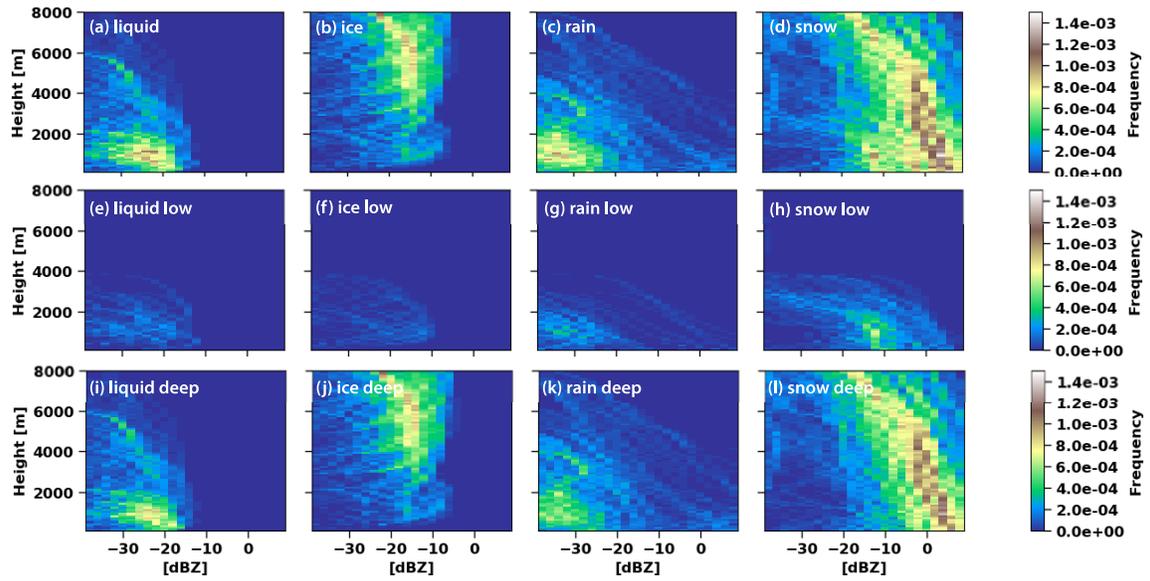
1268  
1269 Fig. 14 (a) CAM6 COSP unattenuated reflectivity, (b) CAM6 COSP attenuated  
1270 reflectivity as viewed from the ground, and (c) AM4 attenuated reflectivity as viewed  
1271 from space, during the 1-15 February 2018 period of CAPRICORN2.



1272

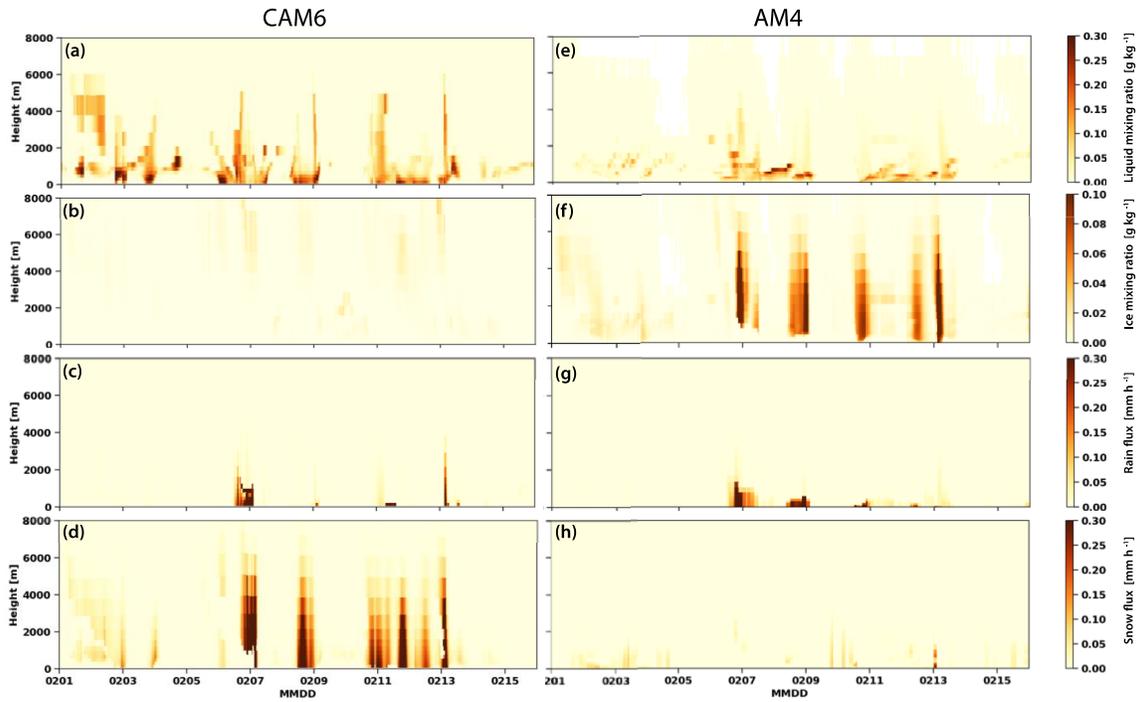
1273 Fig. 15 Contoured frequency by altitude diagrams (CFADs; see text for details)  
 1274 encompassing the entire CAPRICORN2 campaign of (a) W band radar reflectivity  
 1275 observations, (b) CAM6 COSP attenuated reflectivity as viewed from the ground, (c)  
 1276 AM4 COSP attenuated reflectivity as viewed from space. (d)-(f), same as (a)-(c) but for  
 1277 low cloud columns (maximum reflectivity above 4 km  $\leq$  -40 dBZ). (g)-(i), same as (a)-(c)  
 1278 but for deep cloud columns (maximum reflectivity above 4 km  $\geq$  -40 dBZ).

1279



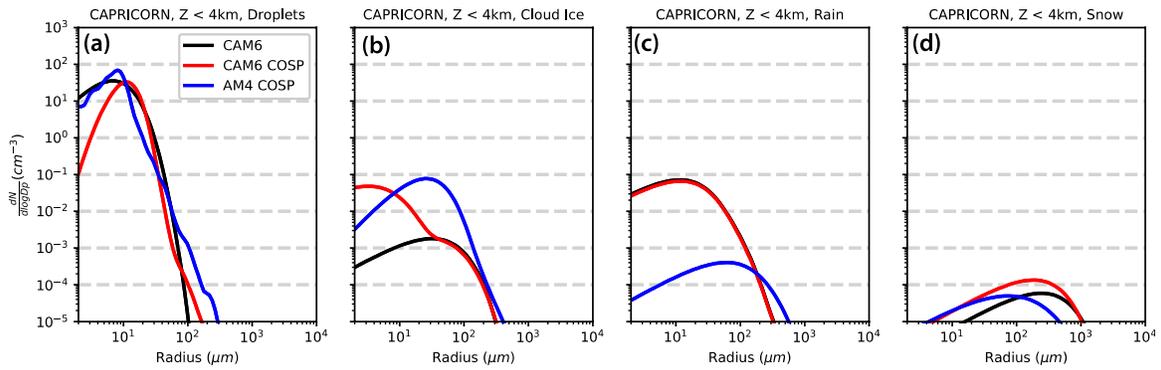
1280

1281 Fig. 16 CFAD for the entire CAPRICORN2 campaign of CAM6 COSP unattenuated  
1282 reflectivity of (a) liquid, (b) ice, (c) rain, and (d) snow. (e)-(h), same as (a)-(d) but for low  
1283 cloud columns. (i)-(l), same as (a)-(d) but for deep cloud columns.



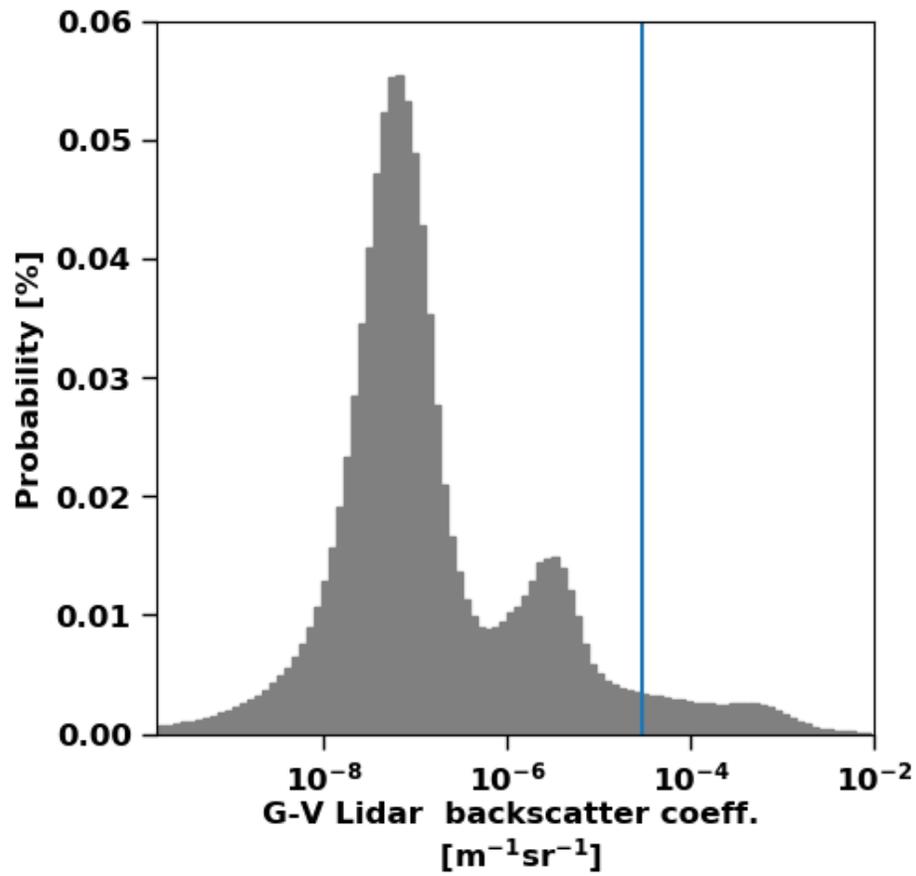
1284

1285 Fig. 17 (a) CAM6 liquid water mixing ratio, (b) CAM6 ice mixing ratio, (c) CAM6 rain  
1286 flux, (d) CAM6 snow flux, (e) AM4 liquid water mixing ratio, (f) AM4 ice mixing ratio,  
1287 (g) AM4 rain flux, and (h) AM4 snow flux during the 1-15 February 2018 period of  
1288 CAPRICORN2.



1289

1290 Fig. 18 Particle size distributions for low clouds in CAM6, CAM6 COSP, and AM4  
1291 COSP during CAPRICORN campaign.



1292

1293 Fig. A1 Probability density function of HSRL backscatter coefficient for 15 flights during  
1294 SOCRATES

1295

1296