Can GCMs represent cloud liquid water path adjustments to aerosol–cloud interactions?

Johannes Mülmenstädt[1], Andrew S. Ackerman[2], Ann M. Fridlind[2], Meng Huang[1], Po-Lun Ma[1], Naser Mahfouz[1], Susanne E. Bauer[2], Susannah M. Burrows[1], Matthew W. Christensen[1], Sudhakar Dipu[3], Andrew Gettelman[1], L. Ruby Leung[1], Florian Tornow[2,4], Johannes Quaas[3], Adam C. Varble[1], Hailong Wang[1], Kai Zhang[1], and Youtong Zheng[5,6]

[1]Atmospheric, Climate and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA
[2]NASA Goddard Institute for Space Studies, New York, NY, USA
[3]Leipzig Institute for Meteorology, Leipzig University, Leipzig, Germany
[4]Columbia University Center for Climate System Research, New York, NY, USA
[5]Atmospheric and Oceanic Science Program, Princeton University, Princeton, NJ, USA
[6]Department of Earth and Atmospheric Science, University of Houston, Houston, TX, USA

Correspondence: J. Mülmenstädt (johannes.muelmenstaedt@pnnl.gov)

Abstract. General circulation models (GCMs), unlike other lines of evidence, indicate that anthropogenic aerosols cause a global-mean increase in cloud liquid water path ($\mathcal{L}$), and thus a negative adjustment to radiative forcing of the climate by aerosol–cloud interactions. In part 1 of this manuscript series, we showed that this is true even in models that reproduce the negative correlation observed in present-day internal variability of $\mathcal{L}$ and cloud droplet number concentration (N_d). We studied several possible confounding mechanisms that could explain the noncausal cloud–aerosol correlations in GCMs and that possibly contaminate observational estimates of radiative adjustments. Here, we perform single-column and full-atmosphere GCM experiments to investigate the causal model-physics mechanisms underlying the model radiative adjustment estimate. We find that both aerosol–cloud interaction mechanisms thought to be operating in real clouds – precipitation suppression and entrainment evaporation enhancement – are active in GCMs and behave qualitatively in agreement with physical process understanding. However, the modeled entrainment enhancement has a negligible global-mean effect. This raises the question whether the GCM estimate is incorrect due to parametric or base-state representation errors, or whether the process understanding gleaned from a limited set of canonical cloud cases is insufficiently representative of the diversity of clouds in the real climate. Regardless, even at limited resolution, the GCM physics appears able to parameterize the small-scale microphysics–turbulence interplay responsible for the entrainment enhancement mechanism. We suggest ways to resolve tension between current and future (storm-resolving) global modeling systems and other lines of evidence in synthesis climate projections.

1 Introduction

Increased aerosol concentration modifies cloud properties by increasing cloud droplet number, which initially makes clouds more reflective. When the aerosol concentration increase is due to an agent external to the climate system, for example, anthropogenic emissions, this cloud-brightening aerosol–cloud interaction (ACI) exerts a negative radiative forcing (RFaci) on

the climate. However, clouds then adjust to the cloud droplet number (N_d) perturbation by changing their liquid water path ($\mathcal{L}$) and cloud coverage; this enhancement or weakening of the instantaneous RFaci is called the radiative adjustment due to $\mathcal{L}$ ($\mathrm{RA}_{\mathcal{L}}$) or cloud fraction ($\mathrm{RA}_{f_c}$).

General circulation models (GCMs) have long disagreed with other lines of evidence on the sign of $\mathrm{RA}_{\mathcal{L}}$, predicting that anthropogenic aerosols increase $\mathcal{L}$ when observational and large-eddy simulation (LES) estimates predict that $\mathcal{L}$ decreases (Bellouin et al., 2020). Recently, Christensen et al. (2023), Varble et al. (2023), and Mülmenstädt et al. (2024) showed that several Coupled Model Intercomparison 6-generation GCMs (Eyring et al., 2016) produce negative correlations between cloud droplet number concentration N_d and liquid water path $\mathcal{L}$ in present-day internal variability. This is welcome news, because the inability of GCMs to match observations was interpreted as GCMs' inability to represent enhanced cloud-top entrainment of dry air at high N_d. Enhanced entrainment is the dominant $\mathrm{RA}_{\mathcal{L}}$ mechanism according to assessments based on multiple lines of evidence.

However, even GCMs that produce negative N_d–$\mathcal{L}$ correlations in the present day still predict an $\mathcal{L}$ increase in response to anthropogenic aerosol emissions. In other words, the causal response of the model climate to secular changes in aerosols has the opposite sign of the correlation in present-day internal variability. This is concerning, for the negative correlation in observations is one pillar on which the sign of $\mathrm{RA}_{\mathcal{L}}$ rests in assessments based on multiple lines of evidence. Part 1 of this manuscript series (Mülmenstädt et al., 2024) offered several hypotheses for confounders that could produce a noncausal negative correlation between N_d and $\mathcal{L}$.

In this manuscript, we return to entrainment-mediated evaporation of clouds as an adjustment mechanism and the question whether this mechanism is represented in GCMs. If so, then GCMs would also agree with the second pillar on which our multiline assessment of $\mathrm{RA}_{\mathcal{L}}$ rests: LES of cloud turbulence–microphysics interactions that shows a causal mechanism by which increased droplet number results in increased entrainment drying of stratocumulus cloud (Ackerman et al., 2004; Bretherton et al., 2007). We will show that a broadly similar causal mechanism appears to exist in GCMs. This is also concerning, because the GCM results suggest that the well-understood and exhaustively LES-modeled $\mathrm{RA}_{\mathcal{L}}$ in subtropical stratocumulus (Sc) clouds may not be representative of the global-mean $\mathrm{RA}_{\mathcal{L}}$.

The results of Mülmenstädt et al. (2024) and this manuscript, taken together, complicate the Bellouin et al. (2020) picture of reduced $\mathcal{L}$ in response to anthropogenic aerosol. It is possible that the known weaknesses of the tools at our disposal (observations, process modeling, and global modeling) are causing us to misunderstand the sign of $\mathrm{RA}_{\mathcal{L}}$. We conclude with recommendations for using the complementary strengths of our tool set to increase the robustness of multiline assessments of ACI adjustments.

2 Data and methods

The hypothesis that our methods are designed to test is that a causal connection exists between N_d and $\mathcal{L}$ in GCM physics that proceeds via enhanced cloud-top entrainment. Thus, we focus on causal and mechanism-denial experiments in single-column

and three-dimensional (3D) atmosphere runs to elucidate the causal link between N_d and $\mathcal{L}$, and on diagnostics of cloud-top entrainment to ascertain that enhanced entrainment is involved in $\mathrm{RA}_{\mathcal{L}}$.

2.1 Models

We use two of the three CMIP6-era models analyzed by Mülmenstädt et al. (2024) that produce an "inverted v"-shaped N_d–$\mathcal{L}$ correlation: the U.S. Department of Energy Exascale Earth System Model (E3SM) and NASA Goddard Institute for Space Studies (GISS) ModelE3. These models have different turbulence schemes; as entrainment-mediated ACI mechanisms must involve at least the turbulence and microphysics parameterizations, it was desirable to include model diversity in this study. ModelE3 and E3SM are used in single-column mode (Sect. 2.3 and 3.2). E3SM is also used in 3D atmosphere runs (Sect. 2.4, 3.1, and 3.3).

In part 1, we used E3SMv1; here we use E3SMv2 (Golaz et al., 2022) instead because it is significantly more efficient at archiving the large, high-frequency fields required for the entrainment diagnostics. E3SMv2 differs from v1 largely in the parametric tuning (Ma et al., 2022) rather than in changes to the physics formulation. The N_d–$\mathcal{L}$ relationship documented in E3SMv1 persists in E3SMv2 for Sc clouds (Fig. S1).

In part 1, we used a ModelE3 parameter tuning derived through machine learning; here we use the default tuning from the ModelE3 development team, which produces a very similar N_d–$\mathcal{L}$ relationship but does not produce the oscillations in surface precipitation at the lowest $N_d$values that resulted from an assertive subgrid-scale multiplier tuning of autoconversion in nonturbulent layers.

2.2 Cloud selection

As the process understanding of entrainment-mediated $\mathrm{RA}_{\mathcal{L}}$ is based chiefly on a small number of canonical subsidence Sc cases, our main focus is on understanding this cloud type in the GCMs, as well. Thus, while the eventual goal is to understand the full spectrum of clouds that occur in the real atmosphere, for now we apply a restrictive set of criteria to maximize the similarity between model clouds and Sc:

- warm (cloud-top temperature warmer than freezing and zero ice water path), overcast (cloud fraction $f > 0.9$) columns,
- in locations where the dynamic–thermodynamic criteria of Medeiros and Stevens (2011) – on pressure vertical velocity ω at 700 and 500 hPa ($\omega_{700\,\mathrm{hPa}} > 10$ hPa d^{-1} and $\omega_{500\,\mathrm{hPa}} > 10$ hPa d^{-1}) and lower tropospheric stability (potential temperature difference $\theta_{700\,\mathrm{hPa}} - \theta_{1000\,\mathrm{hPa}} > 18.55$ K) – are met at least 30% of the time in the annual mean (see part 1),
- during Sc season (northeastern Pacific: Jun–Aug; southeastern Pacific and southeastern Atlantic: Oct–Feb),
- and with an inversion between model levels 10 and 15 from the surface (approximately 750–1400 m).

We also remove cases where the parameterized convection is triggered, as this introduces transport through the planetary boundary layer (PBL) top that is not considered in the entrainment diagnostics (Sect. 2.5). In the single-column experiments

(Sect. 2.3), the convection scheme does not trigger in either ModelE3 or E3SM. In E3SM, the "shallow" convection parameterization is part of the Cloud Layers Unified By Binormals (CLUBB; Golaz et al., 2002; Larson, 2022) cloud macrophysics and turbulence scheme, which does not produce "nonlocal" (i.e., across multiple vertical levels in one time step) transport through the PBL top. The "deep" convection scheme in E3SM (Zhang and Mcfarlane, 1995; Xie et al., 2019) occasionally triggers in the 3D atmosphere run even in the stratocumulus-like conditions we select for; these columns make up 2.0% of the total and are removed from the analysis. The occurrence of deep convection is rare enough that we do not attempt to analyze its effect on time steps after the removed deep-convective time step.

These cloud requirements reduce the complication that the cloud sample may comprise different cloud regimes governed by different ACI mechanisms (Mülmenstädt and Feingold, 2018), ensure validity of the cloud-top entrainment diagnostics, and avoid difficulties in the interpretation of cloud-top entrainment in partly cloudy model columns.

2.3 Single-column model experiments

We use an extensively studied, idealized subsidence Sc-like case specification to construct single-column experiments that are designed to investigate the mechanisms underlying the response of entrainment and $\mathcal{L}$ to N_d in GCM physics.

2.3.1 DYCOMS-II RF02 case description

The initial conditions and forcings for the single-column models (SCMs) follow the specifications of Ackerman et al. (2009) for an intercomparison of lightly drizzling Sc. This setup is based on measurements during research flight 2 (RF02) of the second Dynamics and Chemistry of Marine Stratocumulus (DYCOMS-II) field study off the California coast (Stevens et al., 2003; vanZanten et al., 2005), obtained from horizontally averaging a notably heterogeneous field of somewhat heavily drizzling open cells within barely drizzling closed cells. With regard to the sensitivity of cloud thickness to entrainment in this case, the inversion is sufficiently strong and the overlying air sufficiently dry that it is not close to the "cloud deepening through entrainment" regime of Randall (1984), and thus entrainment is expected to thin the cloud layer, as found in the LES intercomparison. [As an aside, we note that there is a sign error in equation (3) of Ackerman et al. (2009) specifying the total moisture profile above the inversion: the difference in the innermost brackets should be $z_i - z$ rather than $z - z_i$ as written.]

The Wyant et al. (2007) SCM intercomparison study used nearly the same setup as Ackerman et al. (2009), and both studies found that including drizzle and cloud droplet sedimentation generally slowed entrainment and enhanced domain-mean liquid water path among a variety of models. Like the previous LES intercomparison of nocturnal Sc by Stevens (2005), the idealized setup ignored horizontal advective tendencies of cooling and drying associated with the large-scale flow for subtropical Sc decks as well as any solar radiation, consistent with the approximately 5-h aircraft sampling a nocturnal boundary layer air along an approximately Lagrangian trajectory. These and other simplifications such as constant subsidence and turbulent surface fluxes are consistent with the 6-hour simulation duration for the DYCOMS-II RF02 intercomparisons.

2.3.2 GISS ModelE3 SCM description

The GISS ModelE3 SCM is a single-column version of the ModelE3 GCM that includes a number of updates to the column moist physics, as summarized in Cesana et al. (2021) and described in more detail by Cesana et al. (2019); unpublished manuscripts will document the model physics parameterizations more completely and also discuss the machine-learning approach to tuning the atmospheric model. In SCM mode the resolved advection is neglected and vertical advection is treated by multiplying the vertical wind by the local gradient of all prognostic variables, as done for LES with periodic lateral boundary conditions (e.g., Stevens, 2005; Ackerman et al., 2009) to avoid complications associated with representing a divergent flow in a one-dimensional framework. The ModelE3 SCM allows for a number of specified forcings to override the native model parameterizations, which for this case consists of the following: (1) the radiative transfer uses a Beer's Law treatment that computes cloud-top cooling and cloud-base warming from the respective cumulative water paths downward from above and upward from below, (2) geostrophic wind forcing is computed as in the LES framework using a fixed profile of geostrophic wind and the prescribed latitude, and (3) the surface drag is computed using a fixed friction speed.

For the ModelE3 SCM simulations here, we depart from the DYCOMS-II RF02 LES intercomparison specification in two ways. For the sake of simplicity, instead of a bimodal cloud condensation nuclei (CCN) distribution we specify a single lognormal mode of ammonium bisulfate with geometric mean radius 60 nm and geometric standard deviation 1.7. We also extend the duration of the simulations to 24 h to check whether the clouds reach a steady state in each model; the latter half (hours 12.5–24) is not further analyzed.

Appendix A compares the SCM behavior against LES and assesses variations in the SCM setup that differ from those used for the E3SM SCM.

2.3.3 E3SM SCM description

The E3SM SCM is described in Bogenschutz et al. (2020). DYCOMS-II RF02 is part of the standard E3SM SCM case library. The main differences compared with the ModelE3 SCM are as follows. Since E3SMv2, the SCM uses the same vertical advection scheme as the three-dimensional model. The idealizations active in the baseline experiment are prescribed surface heat fluxes, prescribed geostrophic wind, prescribed profile of divergence, and prescribed bimodal aerosol profile, as in the ModelE3 setup. Prescribing the surface wind stress or friction velocity is not supported in the E3SM SCM; during spinup, the SCM stabilizes to $u_* \approx 0.4$ m s^{-1}, substantially higher than the DYCOMS-II RF02 case specification ($u_* = 0.25$ m s^{-1}).

Differences in model configuration in sensitivity experiments and N_d susceptibility scans are described in the discussion of those experiments.

2.4 3D GCM configuration

The models analyzed in part 1 produced $\mathrm{RA}_{\mathcal{L}} < 0$ in the default model configuration, presumably because of precipitation suppression. To disentangle the opposing $\mathrm{RA}_{\mathcal{L}}$ of precipitation suppression and a potential entrainment mechanism ($\mathrm{RA}_{\mathcal{L}} > 0$), we turn off the precipitation suppression in E3SM by setting the exponent on N_d in the autoconversion parameterization to zero,

removing the explicit N_d-dependence of the autoconversion process. To maintain a climate state similar to the default model, we increase the autoconversion scale factor (Mahfouz et al., submitted). This is equivalent to presenting autoconversion with a globally constant $N_d \approx 50$ cm^{-3} and results in present-day top-of-atmosphere radiative flux and cloud radiative effect changes < 1 W m^{-2} compared with the default configuration. The warm-cloud over-ocean mean is $\log \mathcal{L}_{\mathrm{PD}} - \log \mathcal{L}_{\mathrm{PI}} = 2.0 \times 10^{-3}$ in this configuration, compared with 3.2×10^{-2} in the default configuration. This indicates that switching off the precipitation suppression mechanism eliminates the strong negative $\mathrm{RA}_{\mathcal{L}}$ of the default model configuration but does not expose a strong positive $\mathrm{RA}_{\mathcal{L}}$ in its stead. The N_d–$\mathcal{L}$ correlation becomes more negative in Sc clouds when precipitation suppression is turned off (Fig. S2) compared with when precipitation suppression is active.

2.5 Entrainment diagnostics

As the causal ACI mechanism hypothesized to lead to $\mathrm{RA}_{\mathcal{L}} > 0$ is enhanced entrainment with increasing N_d, we make it a focus of this paper to understand how entrainment behaves in the models. To this end, we use an entrainment diagnostic that calculates the mixing between free troposphere (FT) and well-mixed boundary layer as a residual term in the mixed-layer budgets of water and temperature. Note that we do not include budget terms for nonlocal transport through the PBL top by parameterized convection schemes. This limits the applicability of the diagnostics to strictly stratocumulus-topped boundary layers (excluding, e.g., the stratocumulus-to-cumulus transition), which the subsidence-based and stability-based regional selection (Sect. 2.2) is intended to enforce.

Let q_v and q_l be water vapor and liquid mixing ratio, θ and T potential temperature and temperature. In adiabatic expansion and condensation, the total water mixing ratio

$$q_t = q_v + q_l \tag{1}$$

and liquid-water potential temperature θ_l are conserved; we approximate

$$\theta_l = \theta - \frac{L_v}{c_p}\frac{\theta}{T} q_l, \tag{2}$$

with L_v the latent heat of evaporation of water (which, for simplicity, we take as temperature-independent, using its value at 273 K) and c_p the isobaric specific heat of dry air. Budget equations for θ_l, q_t, and total mass, vertically integrated over the PBL, involve fluxes of water, dry air, and heat across the boundaries of the PBL (Lilly, 1968; Stevens, 2002; Caldwell et al., 2005; Kalmus et al., 2014; Mellado, 2017). Crucially for our purposes, this includes the entrainment flux into the boundary layer. We express the budget equations following Kalmus et al. (2014) but modify the notation to highlight the similarity with source and sink terms in a prognostic equation in Lagrangian form:

$$h\frac{\hat{D}(\rho\theta_l)}{\hat{D}t} = -\frac{\Delta F}{c_p} + \frac{L_v \Delta R}{c_p} + \frac{\mathrm{SH}}{c_p} + E_\theta(\theta_l^+ - \langle\theta_l\rangle) \tag{3}$$

$$h\frac{\hat{D}(\rho q_t)}{\hat{D}t} = -\Delta R + \frac{\mathrm{LH}}{L_v} + E_q(q_t^+ - \langle q_t\rangle) \tag{4}$$

$$\rho|_{z=h}\left(\frac{\partial h}{\partial t} + \mathbf{v}\cdot\nabla_H h\right) = E_h - \omega/g, \tag{5}$$

where ρ is the density, ΔF the radiative cooling, ΔR the precipitation mass flux at the surface, LH and SH the latent and sensible heat fluxes at the surface, h the PBL geometric depth, $\mathbf{v}$ the wind vector, $\omega = Dp/Dt$ the large-scale pressure velocity, and g the gravitational acceleration. The operator ∇_H is the horizontal gradient. Quantities with a + superscript [i.e., q_t^+ and θ_l^+ in (3)–(4)] are evaluated just above the inversion. Quantities in angular brackets ($\langle A \rangle$) are mass-weighted vertical averages of a quantity A, evaluated at model-level midpoints k between the lowermost atmosphere level k_{sfc} and the uppermost level below the inversion k_{pbl}:

$$\langle A \rangle = \frac{1}{\langle \rho \rangle h} \sum_{k=k_{\mathrm{sfc}}}^{k_{\mathrm{pbl}}} \rho_k \Delta z_k A_k; \tag{6}$$

the PBL-averaged material derivative of a 3D quantity A is defined as

$$\frac{\hat{D}(\rho A)}{\hat{D}t} = \langle \rho \rangle \frac{\partial \langle A \rangle}{\partial t} + \langle \rho \mathbf{v} \cdot \nabla_H A \rangle + \langle \rho \rangle (A|_{z=h^-} - \langle A \rangle) \mathbf{v} \cdot \nabla_H h. \tag{7}$$

The PBL-averaged material derivative differs from the conventional material derivative in two ways. First, vertical advection is absent, as the motion of the boundary between PBL and free troposphere does not affect the within-PBL average. Second, the final term in (7) accounts for the horizontal advection of the PBL top; the $A|_{z=h^-}$ notation indicates the field is to be evaluated just below the inversion.

Each of the budgets of θ_l, q_t, and h (3)–(5) can be solved for an entrainment mass flux: E_θ, E_q, and E_h, respectively. Physically,

$$E_\theta = E_q = E_h. \tag{8}$$

However, models do not necessarily respect this equality. Therefore, we retain the freedom to diagnose E_θ, E_q, and E_h separately; in the following, we use the degree of equality between these fluxes as a criterion for model fidelity to the physical system.

3 Results

In Sect. 3.1, we describe the effective entrainment in the Sc regime in E3SM according to the entrainment diagnostics introduced in Sect. 2.5. We analyze which properties of the atmospheric column influence the entrainment. To perform an unambiguous demonstration of a causal effect of increased aerosol on entrainment and PBL drying, we then turn to SCM analysis in ModelE and E3SM in Sect. 3.2. We return to the 3D atmosphere in a model configuration without precipitation suppression in Sect. 3.3 to search for evidence of a causal mechanism leading to reduced $\mathcal{L}$ in response to anthropogenic aerosols.

3.1 GCM effective entrainment

In a numerical model, the spatial discretization can potentially alter the behavior of the PBL in a qualitative way. The physical Sc-topped PBL entrains free-tropospheric air by $O(1\ \mathrm{m})$-scale turbulent exchange through a sharp buoyancy barrier (Wood,

2012, and references therein). In the model, the static stability due to thermodynamic jumps across the inversion at PBL top is less localized and weaker due to the finite vertical resolution. Depending on the model resolution, the resolved-scale advection scheme, and the turbulence parameterization, vertical mixing across the poorly resolved inversion may be too strong. This can occur because the stability reported to the turbulence scheme is underestimated or because fluctuations in the resolved-scale vertical velocity mix the airmasses instead of moving the boundary between them. Collectively, we term these behaviors "model artifacts". The problem with model artifacts in mixing is that the effect of such mixing on the PBL temperature and humidity need not have the correct susceptibility to the multitude of anthropogenic perturbations – forcing by and adjustments to both aerosol and greenhouse-gas forcings, as well as feedback mechanisms in response to anthropogenic global warming – that are hypothesized to influence cloud-top entrainment by changing the atmospheric state (N_d, temperature and humidity in the PBL and FT, and FT emissivity).

Therefore, our task is to determine whether the mixing in the model behaves more like physical entrainment or more like artificial mixing. If we calculate E_θ, E_q, and E_h in (3)–(5) as residuals, then they describe the mixing between PBL and FT, including both the entrainment and model artifacts. We then apply three criteria that help us make that determination:

1. In the real atmosphere, $E_\theta = E_q = E_h$ all describe the same entrainment mass flux that comes about due to turbulent processes at the boundary-layer top. In a numerical model, however, equality of the entrainment fluxes is not a given. For one thing, models treat θ_l and q_t differently, for example, to ensure nonnegative-definite q_t. For another, the length scales at which entrainment occurs reach below 1 m, far beyond the resolved dynamics of most types of models. Mixing between the boundary layer and FT in a model, therefore, results from a combination of resolved advection and parameterizations. Having multiple independent measures of the entrainment mass flux affords us the ability to ask both whether the model-diagnosed entrainment estimates are consistent and whether they are physical. Consistent fluxes are highly correlated, with E_q, E_θ, and E_h close to a 1:1 regression slope. (In an Eulerian model, E_h is difficult to diagnose when the advective tendency of h over a model time step is small compared to the vertical resolution, which is the case in GCMs. We restrict our analysis to E_q and E_θ.)

2. Even if the diagnosed entrainment fluxes are consistent, however, they can still be unphysical. That is, the mass flux could be detraining air out of the boundary layer instead of entraining into the boundary layer.

3. Finally, the dependence of the entrainment flux on atmospheric conditions can indicate that the wrong mechanisms are at work in the model. For example, a strong dependence of entrainment on the FT vertical velocity would indicate overly strong vertical advection through the capping inversion.

Thus, we propose three measures of the realism of the entrainment representation in a model: the joint distribution of E_θ and E_q, the sign of the mass flux, and the dependence of the entrainment flux on the atmospheric state.

Under Sc conditions (as defined in Sec. 2.2), E3SM produces vertical profiles of q_t and θ_l consistent with a fairly well-mixed PBL capped by a fairly sharp thermodynamic jump. Figure 1 shows composite vertical profiles stratified by PBL depth.

The effective entrainment qualitatively agrees very well with physical understanding of the Sc-topped PBL. The fluxes derived from the separate budgets agree well with each other, yielding a close relationship with slope near 1 (see joint probability

in Fig. 2). Furthermore, the sign of the fluxes is consistent with physical entrainment from the FT into the PBL ($E > 0$; see the marginal cumulative distribution functions in Fig. 2) rather than showing a distribution including both positive and negative values, which would be consistent with numerical diffusion. The instantaneous entrainment also qualitatively responds in the expected way to instantaneous variability (as opposed to climatological spatial variability, seasonal temporal variability, etc.) in properties of the atmospheric column (Fig. 3). Entrainment increases with surface heat fluxes and cloud-top radiative cooling, consistent with increased turbulence production leading to increased entrainment; decreases with the magnitude of the θ_l jump at the inversion, consistent with a stronger buoyancy barrier suppressing entrainment (the negative q_t jump, which reduces the buoyancy barrier, is strongly correlated with the positive θ_l jump); and is independent of the instantaneous grid-scale vertical velocity, consistent with large-scale subsidence moving the boundary between airmasses (i.e., the FT and the PBL) rather than mixing them.

In summary, the entrainment behavior of the GCM, according to the criteria tested for, appears free of numerical artifacts that are due to coarse model resolution. While not exhaustive (e.g., the effect of mixing on the model levels above the inversion is not examined), these entrainment diagnostics allow us to compare the GCM physics to the foundational LES modeling studies on Sc entrainment (e.g., Ackerman et al., 2004; Bretherton et al., 2007; Ackerman et al., 2009), which diagnose entrainment from the entrainment velocity, i.e., the PBL mass budget. We next focus on the effects of anthropogenic perturbations on the modeled entrainment and on how the parameterized model physics affects those entrainment responses.

3.2 Entrainment ACI mechanism in single-column runs

Figure 3 shows that entrainment depends on numerous properties of the atmospheric column. This creates a bewildering web of possible causal and covariability effects by which entrainment internal variability could be correlated with aerosol internal variability in 3D atmosphere runs. SCM runs, in contrast, provide a clean way to diagnose cause and effect in GCM column physics, which is a reason they are widely used during model development. First, the SCM allows us to hold any combination of boundary conditions on a single column fixed. Thus, we can switch off any effects mediated by the grid-scale horizontal circulation. These effects includes synoptic-scale confounders of the type discussed by Mülmenstädt et al. (2024). Second, and relatedly, we are free to explore the effects of model physics on ACI without having to retune the model to global-mean energy balance. Thus, we avoid the difficult problem whether to attribute changes in model behavior to the physics changes under investigation versus the nuisance changes required to restore energy balance that might also affect the ACI behavior (e.g., Golaz et al., 2011; Mülmenstädt et al., 2020, 2021).

A welcome side effect (and a raison d'être) of SCM use is that the experiment setup closely matches the LES runs that inform so much of our process understanding. Like Ackerman et al. (2004), Bretherton et al. (2007), and Hoffmann et al. (2020), we can focus on well-understood subtropical subsidence-region Sc and vary one boundary condition – the aerosol concentration – at a time (and, if desired, one model-physics mechanism at a time). This does not address the question whether the model results are representative of the global-mean effective radiative forcing – that is best addressed with global runs – but it answers the question whether LES and GCM column physics respond similarly to perturbations around an as-near-as-possible identical base state.

The E3SM SCM uses the two-mode prescribed aerosol concentration profile specified for RF02 (Wyant et al., 2007), while the ModelE3 SCM uses a single accumulation mode aerosol size distribution as described above. We then modify the amplitude of this profile to elucidate the causal effect of N_d change on $\mathcal{L}$. In the ModelE3 SCM, we scan the aerosol number concentration $N_a = \{20, 30, 40, 60, 80, 120, 160, 320, 640\}$ cm^{-3}. In the E3SM SCM, we scale the prescribed aerosol concentration up and down by a factor of 8: $N_a = N_{a_0} \times \{1/8, 1/4, 1/2, 1, 2, 4, 8\}$. Prescribing aerosol eliminates ACI mechanisms in which clouds affect the aerosol state, such as potential effects of aerosol scavenging on the N_d–$\mathcal{L}$ relationship (McCoy et al., 2020). This simplifies the attribution of $\mathcal{L}$ responses to N_d perturbations by removing one class of processes from consideration.

Figures 4 and 5 show the ModelE3 and E3SM SCM time series. In both models, the PBL deepens, indicating entrainment in excess of the subsidence rate; the PBL in both models also deepen a similar amount when native longwave radiative transfer is used (see Appendix). With the exception of a short duration before and after steps in the discretized PBL depth in E3SM, both models maintain an overcast cloud; the loss of cloud cover in E3SM's SCM when the PBL top jumps by a model level is clearly a model artifact, so these periods are excluded from further analysis by requiring $f > 0.9$, as in the 3D model analysis both here and in part 1. Furthermore, the discrete PBL depth increases are associated with discontinuities in the entrainment diagnostics: a dependence of E on how long the top model level of the PBL has been subject to entrainment in E3SM, and a reversion to a constant E once the PBL has deepened in ModelE3. We may be mitigating the E3SM artifact by averaging over two full deepening cycles, effectively averaging over the dependence of E on position in the deepening cycle. We attempt to mitigate the ModelE3 artifact by only averaging E until the first PBL deepening occurs. This choice is subjective, and we note that the other obvious choice, which is to average the entrainment over the same time period (hours 2–12) as the other variables, produces no clear relationship between E and N_d. However, the problem with this averaging method is that earlier deepening in the more entraining cases means that these cases spend more time in the apparent low-entrainment state that characterizes the entrainment-diagnostics artifact, which most strongly lowers the average apparent entrainment in the highest-entrainment cases. On balance, it appears to us that the averaging method we choose better summarizes the behavior of ModelE3, namely that the greater magnitude of the entrainment flux with increasing N_d is consistent with the earlier deepening.

As expected, varying the aerosol concentration strongly affects the droplet concentration. From this response, the two SCMs then diverge in the details of their behavior, but they reach the same behavioral endpoint at sufficiently large N_d: enhanced entrainment leading to a loss of $\mathcal{L}$ as N_d increases. In ModelE3, $\mathcal{L}$ after spinup starts out with a monotonically increasing N_d dependence; after spinup, the low-N_d runs experience an increase in $\mathcal{L}$, while the high-N_d runs experience a decrease, leading to a time-average $\mathcal{L}$ that first increases with N_d and then decreases (Figs. 4 and 6). The increasingly negative $\mathcal{L}$ tendency as a function of N_d accompanies an increasingly strong entrainment warming and drying (Fig. 6). The increase in entrainment with increasing N_d competes with a decrease in precipitation (Fig. 4). ModelE3 with native longwave radiation, however, maintains three times greater $\mathcal{L}$ after 24 h duration (see Appendix).

In E3SM, $\mathcal{L}$ after spinup has a monotonically decreasing relationship with N_d, which remains true at each point in time throughout the runs. As all runs experience fairly rapid $\mathcal{L}$ loss with time, time averages only show weak dependence on N_d (Fig. S3). We can instead quantify susceptibilities by scanning across the different aerosol experiments at each time step; this is shown for $\mathcal{L}$ in Fig. 7 and for E in Fig. 8. Precipitation in E3SM largely ceases after the first time step, even though drizzle

was measured in the DYCOMS-II RF02 observations. In ModelE3 with native LW radiation, precipitation experiences sharp peaks with a periodicity similar to PBL depth increases.

Whether the effect of varying the aerosol concentration on $\mathcal{L}$ is expected depends on our Bayesian prior. The agreement on $\mathrm{RA}_{\mathcal{L}} < 0$ across CMIP5-era GCMs (Gryspeerdt et al., 2020), which persists in the newer-generation models examined in part 1, and the absence of enhanced entrainment physics in some GCMs (e.g., Salzmann et al., 2010) have led to the notion that GCMs may be structurally incapable of representing the enhanced entrainment mechanism known from LES (e.g., Zhou and Penner, 2017). This thinking is in tension with a long line of parameterization work (e.g., Randall et al., 1985; Lock et al., 2000; Bretherton and Park, 2008; Guo et al., 2011; Karset et al., 2020) indicating that it may well be possible to represent this mechanism in GCMs, either through direct parameterization or as an emergent behavior. If our expectation is based on the LES-based process understanding, then we would predict the causal effect of N_d to be a decrease in $\mathcal{L}$. The surprising result is that the SCM sides with the LES (see Appendix) rather than the 3D GCM with which the SCM shares its model physics. Recall that the E3SM 3D GCM run, in contrast to the E3SM SCM run, showed negligible response of climatological $\mathcal{L}$ to the anthropogenic N_d increase (Sect. 2.4).

While the SCM behavior is consistent with our mechanistic understanding of entrainment-mediated drying, we need to point out several caveats. First, the details of what "entrainment-mediated drying" entails are different in the two models, as discussed above, and we will find in Sec. 3.3 that the behavior in the 3D E3SM run is more consistent with the ModelE3 SCM than with the E3SM SCM. Second, the E3SM SCM entrainment fluxes reach the equivalent of several centimeters per second entrainment velocity, significantly stronger than LES or the ModelE3 SCM produce for this case, and stronger than the E3SM 3D run produces for Sc on average. Third, if the E3SM physics had not serendipitously produced very low precipitation rates, the decrease in $\mathcal{L}$ with increasing N_d would probably have been overwhelmed by the precipitation suppression signal. None of these caveats negates the finding that the GCM physics appears capable of producing entrainment-mediated $\mathcal{L}$ loss qualitatively consistent with LES findings, and significant intermodel diversity is to be expected in SCM studies (Zhu et al., 2005; Wyant et al., 2007). They do, however, indicate that there is ample further process investigation to be performed in future work.

We conduct several additional E3SM SCM experiments with perturbed physics. These experiments further test that the causal effect of aerosols on $\mathcal{L}$ in SCM mode not only has the same sign as in LES but proceeds via the same physical mechanisms and show that the entrainment-mediated drying can be tuned to agree quantitatively with LES.

Sedimentation–entrainment feedback is the source of entrainment enhancement Size-dependent sedimentation is one of the processes by which higher-N_d clouds lose liquid relative to lower-N_d clouds in LES (Bretherton et al., 2007) under sufficiently dry overlying air (Ackerman et al., 2004). The Gettelman et al. (2015) microphysics parameterizes size-dependent sedimentation. Guo et al. (2011) showed through process denial experiments that $\mathcal{L}$ loss only occurs when this process is included in the SCM version of the Geophysical Fluid Dynamics Laboratory (GFDL) AM3 model. This is true in the E3SM SCM as well; Fig. 9 shows that $\partial \log \mathcal{L} / \partial \log N_d$, which is negative when the size-dependent sedimentation is active, becomes ≈ 0 when we switch off the sedimentation flux and thus its size dependence.

Parameter tuning can move the GCMs toward quantitative agreement with LES on susceptibility The ModelE3 SCM closely replicates LES of the RF02 case (Appendix A). While the E3SM SCM behaves significantly differently than LES in that it is nonprecipitating for most of the run, it does appear that its entrainment-mediated $\mathcal{L}$ susceptibility can be moved closer to LES estimates by appropriate parameter choices. Figure 9 also shows that increasing the size-dependent sedimentation by a factor of 2 makes $\partial \log \mathcal{L} / \partial \log N_d$ more negative. Quantitatively, this brings the E3SM SCM closer to quantitative agreement with LES ($-0.35 \leq \partial \log \mathcal{L} / \partial \log N_d \leq -0.22$; Ackerman et al., 2004, supplementary table 1). It may be possible to achieve quantitative agreement by combining the sedimentation tuning factor with similar tuning factors in the turbulence parameterization. This would not be an outlandish model tuning, considering that it may be taking the role of an "enhancement factor" (Covert et al., 2022) compensating for the coarse vertical discretization $O(100\ \mathrm{m})$ compared to the process scale $O(1\ \mathrm{m})$.

The liquid water path reduction could also have an important shortwave absorption component The DYCOMS-II RF02 SCM specification calls for a nocturnal simulation, i.e., without shortwave radiative effects. In the E3SM SCM, we perform a sensitivity test with shortwave radiative effects in which the sun rises at 12 h (not shown); this simulation is less entraining overall, deepening only once, and is only able to sustain its cloud cover for ≈ 18 h. During the daytime portion (hours 12–18), this configuration's $\partial \log \mathcal{L} / \partial \log N_d$ is more negative than the LW-only simulation's. Absorption of shortwave at cloud top exerts a warming effect that increases with N_d (Stephens, 1978; Hoffmann et al., 2020). The behavior of the SCM run with shortwave radiation is consistent with this N_d-dependent cloud-top heating. Cloud-top heating counteracts longwave cloud-top cooling, reducing the entrainment. Shortwave heating increases with increasing N_d, decreasing $\mathcal{L}$.

3.3 Entrainment ACI mechanism in 3D model runs

If the entrainment-mediated adjustment of $\mathcal{L}$ is evident in the SCMs, what becomes of it in the full 3D model atmosphere? From Mülmenstädt et al. (2024), we know that $\mathrm{RA}_{\mathcal{L}} < 0$, that is, $\mathcal{L}$ under present-day (PD) emissions is greater than $\mathcal{L}$ under preindustrial (PI) emissions, opposite in sign to what is expected from the entrainment-mediated mechanisms and from the SCM. To understand why this happens, we use diagnostics targeted at entrainment mechanisms and perform model experiments designed to isolate entrainment mechanisms.

3.3.1 Indications of entrainment mechanisms in present-day correlations

Figure 10a shows cloud-top entrainment into the Sc PBL as a function of N_d and $\mathcal{L}$. Two things are readily apparent in this figure. First, entrainment has a strong dependence on $\mathcal{L}$. Entrainment is expected to increase with $\mathcal{L}$ based on the ability of the cloud to generate turbulence, which depends on the availability of liquid water for evaporation and cloud-top radiative cooling. The $\mathcal{L}$ values at which entrainment turns on are in reasonable agreement with recent LES (Hoffmann et al., 2020) and observational (Zhang and Feingold, 2023) results. Zhang and Feingold (2023), reporting on MODIS retrievals, state 50 g m^{-2} as the threshold for efficient radiative cooling. Hoffmann et al. (2020), reporting on LES experiments, are less explicit, but,

depending on choice of entrainment parameterization and for their $\mathrm{LWP}_\infty = 60$ g m^{-2} simulation, their Fig. 4 shows a sharp turn-on that saturates between approximately 20 and 40 g m^{-2}.

Second, at a given $\mathcal{L}$, entrainment increases with N_d. This is shown quantitatively in Fig. 11: in all three Sc regions (see Sect. 2.2), the entrainment susceptibility $\partial \log E / \partial \log N_d|_{\mathcal{L}} > 0$ except at low $\mathcal{L}$, and the cloud water loss increases (the Eulerian tendency $\partial \mathcal{L}/\partial t$ becomes more negative) with increasing N_d.

The main conclusion from these plots is that the model produces greater entrainment in response to higher N_d in Sc clouds with high enough $\mathcal{L}$ to support strong entrainment. In other words, there appears to be mechanistic agreement between the model physics and process understanding of $\mathrm{RA}_{\mathcal{L}}$ via entrainment enhanced by increased droplet number. Further support for this conclusion comes from the instantaneous $\mathcal{L}$ tendency $\partial \mathcal{L}/\partial t$. The regression slope $\partial^2 \mathcal{L}/\partial N_d \partial t|_{\mathcal{L}}$ is predominantly negative except at low $\mathcal{L}$. In other words, clouds with a positive entrainment susceptibility also exhibit a negative liquid-water tendency, the magnitude of which increases with N_d, confirming that there is a relationship between entrainment susceptibility and cloud water loss (presumably to drying).

There is reason to be cautious, however. We have tested whether the known negative N_d–$\mathcal{L}$ correlation occurs in conjunction with positive N_d–E and negative N_d–$\partial \mathcal{L}/\partial t$ correlations, consistent with an entrainment drying adjustment to an N_d increase. This peels away one layer of confounding between N_d and $\mathcal{L}$, increasing our confidence that there is a mechanistic link between N_d, E, and $\mathcal{L}$. However, it is possible that the relationships of E and $\partial \mathcal{L}/\partial t$ with N_d are themselves confounded, just as the regression between N_d and $\mathcal{L}$ was found to be a result of covariability by Mülmenstädt et al. (2024).

3.3.2 Absence of entrainment-mediated adjustment in PI and PD emissions experiments

In the Sc regime, there are hints of behavior in accordance with process understanding (e.g., Randall, 1984; Ackerman et al., 2004): $\mathcal{L}$ decreases in response to the anthropogenic N_d increase when the relative humidity (RH) in the FT (diagnosed from the first model level above the inversion) is lowest, as shown in Fig. 12. However, there are multiple reasons these results should be treated with caution until they can be confirmed in longer model runs. First, the dependence on FT RH is not robust across Sc regions. While northeast Pacific (NEP) and southeast Pacific (SEP) Sc regions show negative $\mathcal{L}$ susceptibility to anthropogenic N_d at low FT RH and positive $\mathcal{L}$ susceptibility at high FT RH (with $\mathcal{L}$ decrease overall), the southeast Atlantic (SEA) Sc region shows no clear pattern. Furthermore, the NEP and SEP behavior requires us to select only the completely overcast (cloud fraction $f = 1$, rather than the default $f > 0.9$ requirement we use elsewhere; this reduces the data sample from approximately 3.7×10^5 to 1.2×10^5 columns). Otherwise, the $\mathcal{L}$ susceptibility has no clear sign or FT RH dependence across regions. A final puzzling observation is that the entrainment mass flux, entrainment temperature flux, and entrainment moisture flux are all virtually unchanged between PD and PI emissions. On balance, the conservative interpretation of these results is that any potential $\mathcal{L}$ reduction signal in response to anthropogenic aerosol is small enough to require far longer model runs to detect.

In the global mean, $\mathcal{L}$ is virtually unchanged when emissions are changed from PI to PD in our E3SM configuration with deactivated precipitation suppression. (The reader may recall from Sect. 2.4 that the global-mean warm, overcast cloud $\Delta \log \mathcal{L} = 2.0 \times 10^{-3}$.) Thus, even if an entrainment-drying ACI mechanism is represented in the model (as the evidence from

the SCM experiments in Sect. 3.2 and the PD statistics in Sect. 3.3.1 suggests), the model considers that mechanism's global effect to be negligible.

One possible explanation why the positive susceptibility of E to N_d seen in PD internal variability does not lead to a decrease in $\mathcal{L}$ is that entrainment susceptibility may beget its own demise (Zhu et al., 2005; Wood, 2012). (We reiterate, as throughout, that another possible explanation for relationships seen in internal variability is confounding.) It is true that E appears to increase with N_d at a given $\mathcal{L}$ (Fig. 10a). However, $\mathcal{L}$ is not fixed during the temporal evolution of a cloud; as long as the cloud remains surface-coupled, closed-cell Sc, increased entrainment leads to loss of $\mathcal{L}$. But at lower $\mathcal{L}$, entrainment is weaker. Eventually, entrainment may even decrease $\mathcal{L}$ to a low enough value that the cloud is protected from further entrainment drying (Hoffmann et al., 2020; Zhang et al., 2022). Cloud aggregate statistics are consistent with this interpretation, where the E susceptibility to N_d at fixed $\mathcal{L}$ is positive (Fig. 10a), but the overall susceptibility of E on N_d still becomes negative at sufficiently high N_d (Fig. 10b) due to the strong negative correlation between $\mathcal{L}$ and N_d. If such a negative feedback mechanism is at play, it would be an example of buffering in the cloud system (Stevens and Feingold, 2009): an initial cloud loss process being shut off by the change in cloud state due to that process.

4 Discussion, conclusions, and recommendations

We have documented two surprising behaviors from GCMs. The first is that GCMs can produce negative N_d–$\mathcal{L}$ PD correlations that do not predict $\mathrm{RA}_{\mathcal{L}}$ (Mülmenstädt et al., 2024). In terms of mechanistic understanding, the simplest explanation is that the correlation is due to confounding rather than a causal relationship involving the entrainment-mediated mechanisms suggested by process scale modeling.

This is where models had a second surprise in store: there is actually a causal negative relationship between aerosol and $\mathcal{L}$. The evidence for this causal relationship comes from SCM studies, where, like Guo et al. (2011), we find that increased aerosol, while holding all other boundary conditions fixed, leads to liquid-water loss at sufficiently high N_d. Furthermore, this loss appears to be due to increased entrainment, or at least it occurs in conjunction with increased entrainment when the aerosol boundary condition is increased. At face value, this would seem to indicate excellent mechanistic agreement with LES-based process understanding that enhanced entrainment drying reduces $\mathcal{L}$. Three-dimensional atmosphere runs, too, show evidence for entrainment-mediated liquid-water loss in correlations between entrainment and N_d.

However, like Karset et al. (2020), we find that a secular N_d increase caused by anthropogenic emissions leads, at best, to a very weak decrease in $\mathcal{L}$, unlike what would be expected from the SCM idealized case study or the relationships found in PD internal variability (increased E, increasingly negative $\partial\mathcal{L}/\partial t$ when N_d increases). This may be a manifestation of buffering of the cloud system against perturbations; a candidate for the buffering mechanism is that enhanced entrainment leads to sufficient liquid-water loss to shut off entrainment driven by cloud-top radiative cooling, protecting the clouds from further liquid loss.

Summarizing the findings from parts 1 and 2 of this manuscript series, we come to the following conclusions. First, negative relationships between N_d and $\mathcal{L}$ observed in PD internal variability are not necessarily indicative of a causal reduction in $\mathcal{L}$, and thus not necessarily predictive of decreased $\mathcal{L}$ when N_d increases due to anthropogenic emissions in GCMs. Second, causal

negative relationships between N_d and $\mathcal{L}$ in LES are not necessarily representative of the more diverse ensemble of clouds in the global-mean $\mathrm{RA}_{\mathcal{L}}$. Thus, the disagreement on the sign of $\mathrm{RA}_{\mathcal{L}}$ between global models and other lines of evidence (Bellouin et al., 2020) may not be solely due to a deficiency in the GCM physics; it could also be due to known deficiencies in the other lines of evidence.

Casting doubt on whether we even know the sign of $\mathrm{RA}_{\mathcal{L}}$ is a highly unsatisfactory state of affairs. Answering the following six questions would provide a potential remedy:

What complexity is required? Many different processes are at play, and it is not clear which ones are represented in the models studied here, either through parameterization or emerging from the interplay of physics and dynamics. There is certainly value in (and models are suited to) studying how the climate response depends on the processes included the model. But trying to include all known or hypothesized processes could fall into the trap of amassing a zoo of "*n*th indirect effects" (Stevens and Feingold, 2009; Mülmenstädt and Feingold, 2018). Instead of overelaborating the model with redundant and competing parameterizations (Proske et al., 2023), the approach leading to the lowest climate projection uncertainty may lie in finding the minimal set of parameterizations that allow the model to reproduce physical process understanding of the sensitivities that matter for the climate problem: sensitivities to those boundary conditions that change with aerosol and greenhouse-gas ERF or with global warming. Evident qualitative differences in the E3SM and ModelE3 SCM behaviors compared with one LES and the challenges of well-constraining LES with observations are reminders that the representation of basic microphysical and turbulent processes still afford ample opportunity for tighter constraint.

What resolution is required? GCM resolution can offer, at best, a cartoon version of the mechanisms at play in real clouds. But "cartoon" does not have to be a pejorative; it is the simplest representation of reality that can convey the author's intent (wit, satire, heuristic simplification, or, when applied metaphorically to models, predictive skill for a different climate state). As noted in the previous paragraph, there is value in simplicity. There is a trade-off between resolution and simplicity, however: the coarser the resolution, the greater the reliance on the parameterized physics, and the longer the list of phenomena that need to be parameterized. Can the climate effect of entrainment be adequately captured by "bulk" entrainment throughout the GCM grid box, or is the mesoscale variability, in the form of the cloud-top circulation engulfing free-tropospheric air (e.g., Yamaguchi and Randall, 2012; Zhou and Bretherton, 2019), essential for a correct projection of the entrainment response to anthropogenic perturbations? (And, in that case, are km-scale "storm-resolving" global models able to represent these features by beginning to resolve the mesoscale dynamics of large Sc cells?) How far this trade-off between resolution and parameterization can be pushed determines the minimal resolution required for reliable climate projections. Terai et al. (2020) and other studies already provide hints at the answer. An important additional piece of information that can be obtained from the entrainment diagnostics presented here is how entrainment behavior changes, qualitatively and quantitatively, as model resolution is coarsened from LES (or, ideally, direct numerical simulation of the cloud-top turbulence; Mellado et al., 2018) to global storm-resolving models to 10–100 km-scale GCMs.

How do base-state and process errors affect modeled climate responses? Our results show that entrainment and its susceptibility are strong functions of $\mathcal{L}$ and N_d. The climate response may also be a function of the model's FT RH, which appears biased high in E3SMv2. Dependence on base state (Christensen et al., 2023; Varble et al., 2023) and competing processes (Mülmenstädt et al., 2020, 2021) in models necessitates careful evaluation of the base state N_d and $\mathcal{L}$, paying close attention to issues of definition and aggregation (Elsaesser et al., 2017; Feingold et al., 2022; Varble et al., 2023). Better constraints on the base state alone can run into "equifinality" (von Bertalanffy, 1950; Beven and Freer, 2001; Lee et al., 2016; Regayre et al., 2018; Mülmenstädt and Feingold, 2018) problems that negate a direct reduction in climate projection uncertainty (Lee et al., 2016; Regayre et al., 2018; Mülmenstädt et al., 2020, 2021; Zelinka et al., 2022). In the case of entrainment, however, the apparent strong dependence of the process representation on base-state errors may yield a significant payoff in tighter constraints on climate projections when the base state is improved.

What observational constraints on the entrainment process are available? Along with vital advances in teasing causality out of observations of cloud N_d and $\mathcal{L}$ (Fons et al., 2023), the biggest step forward along the observational track would be better constraints on the entrainment process itself. One possibility may be to diagnose subadiabaticity (Merk et al., 2016; Varble et al., 2023) as an indicator of the cloud liquid loss. This would only provide a time-integrated measure of the loss processes, and as such would require disentangling entrainment drying from precipitation.

How representative are susceptibilities derived in small ensembles of individual cases? Given the difficulty of placing observational constraints on entrainment, the most convincing evidence for entrainment-mediated $\mathrm{RA}_{\mathcal{L}}$ continues to come from LES studies. Large ensembles of LES cases (e.g., Gustafson Jr. et al., 2020; Glassmeier et al., 2019) are vital to provide resilience against the possibility that the well-studied, often idealized canonical subsidence Sc conditions may not be representative of the global-mean role that cloud-top entrainment plays in ERFaci. These LES ensembles will be particularly valuable if they span the initial-value and boundary-value problem aspects of the climate response (i.e., sample the vast variability of meteorology encountered by Sc clouds in the climate), and if they provide cloud lifecycle evolution (Kazil et al., 2021) that can be validated against cloud lifecycle observations (Christensen et al., 2020) sufficiently to ensure LES adequacy for purpose, given differing results in multi-LES studies (e.g., Ackerman et al., 2009). The same point on the importance of large ensembles holds for SCM studies: the differences between E3SM and ModelE SCM of the DYCOMS-II RF02 case, as well as the differences between the high-entrainment E3SM SCM and moderate-entrainment E3SM 3D runs, illustrate the need for a set of SCM test cases that better approximate the diversity of meteorological conditions encountered in the climate. A way forward would be to perform the suite of SCM causal aerosol perturbation experiments and mechanism denial experiments from Sect. 3.2 on an ensemble of single-column cloud cases from a 3D run using the SCM ability to "replay" the forcing of the column by the 3D atmosphere (Bogenschutz et al., 2020).

Was precipitation suppression the bigger problem all along? According to our results, and consistent with Karset et al. (2020), cloud-top entrainment, even when represented in the model physics, only appears to play a small role in the

global-mean $\mathrm{RA}_{\mathcal{L}}$. If this GCM finding reflects reality, focusing on the precipitation-mediated component of $\mathrm{RA}_{\mathcal{L}}$ takes on renewed importance.

Recognizing that the ACI climate problem is at its core a multiscale physics problem is crucial, as is recognizing that no single line of evidence is capable of putting our knowledge of $\mathrm{RA}_{\mathcal{L}}$ on solid footing (Mülmenstädt and Feingold, 2018). The above research questions are a sketch of a multiscale modeling and observations roadmap. Simultaneously, by accounting for the multiscale nature of the problem, they would put us on a path of reliable climate projections beyond the global energy budget (e.g., projecting regional hydrologic extremes due to the spatial heterogeneity of ERFaci) by ensuring that global modeling systems correctly represent both the spatial pattern of ERF and the response of the circulation at all scales to this forcing (Mülmenstädt and Wilcox, 2021).

Appendix A: Tracing DYCOMS-II RF02 model behavior from LES to SCM

The SCM setup here is based on the Ackerman et al. (2009) intercomparison of LESs, which is in turn based on airborne observations of a nocturnal marine Sc deck during the DYCOMS-II project (Stevens et al., 2003; vanZanten et al., 2005) following an approximate Lagrangian trajectory over 5 hours (to paraphrase the description of Wyant et al., 2007). The purpose of this appendix is to connect LES of the case to the SCM setups and results in this study.

A representative LES in that study was the Distributed Hydrodynamic Aerosol and Radiative Modeling Application (DHARMA) model, here run with two-moment cloud microphysics (Tornow et al., 2021), and with a vertical grid spacing of $\delta z = 5$ m to 200 m above the original inversion to avoid a positive feedback between entrainment and grid spacing that arises on the Ackerman et al. (2009) specified grid, which was designed to accomodate simulations of duration 6 h instead of the 24 h used here.

The ModelE3 SCM is run here following the specifications of Ackerman et al. (2009), with one departure being that aerosol are treated as a monomodal lognormal distribution with a fixed number mixing ratio (corresponding to 60 cm^{-3} at 900 hPa and 10°C) instead of the bimodal size distribution specified by Ackerman et al. (2009), which produces comparable N_d values to the LES, as seen in Fig. A1.

The LES and ModelE3 SCM model setups also both depart from the Ackerman et al. (2009) specificaton of cloud-water sedimention and instead use the treatment in their (similar) native microphysics schemes of cloud water sedimentation [assuming a gamma distribution with a relative dispersion of 0.3, per Geoffroy et al. (2010) in the SCM, and with a relative dispersion of 0.2 in the LES].

Given that this case was used for the development and default tuning of the ModelE3 SCM, it is not a surprise that the SCM results match the LES reasonably well, with the greatest differences being (1) modestly slower entrainment and, thus, deepening of the marine boundary layer (MBL) in the SCM, and (2) about a factor of two less drizzle reaching the surface for most of the duration. While the formulations of the SCM and LES are different, and such differences are to be expected, we note that a narrowing of the assumed droplet size distribution to match that in the LES has little impact and does not deepen the MBL more over the 24-h duration (not shown). We also note that the stronger drizzle for the LES is not explained by its

assumption of a narrower raindrop size distribution, which on its own would instead be expected to result in weaker drizzle at the surface.

As briefly noted in the main text, the SCM setup used for the E3SM further departs from the Ackerman et al. (2009) specification and the ModelE3 SCM setup here in a number of ways, among them: (1) it did not use the specified surface stress, (2) it did not adopt the specified Beer's Law parameterization of longwave flux divergence, and (3) it did not apply the local subsidence rate to vertical gradients using first-order upwinding to avoid complications with divergent flow, but instead treated the specified divergence using the dynamic core. While we are unable to even begin to match departure (3), we are able to consider departures (1) and (2) with the ModelE3 SCM. For (1), we adopted the equilibrated surface stress from the E3SM results, which has has very little impact on the results (not shown). For (2) we used the ModelE3 native longwave radiation scheme, using the ModelE3 SCM standard machinery to patch in the McClatchey (1972) standard atmosphere above the 1.5-km top of the initial sounding provided by Ackerman et al. (2009). As seen in Fig. A1, doing so results in appreciably faster entrainment, which better matches the E3SM SCM results (Fig. 2), which also deepens by about 350 m over the 24-h duration.

The inverted-v relationship between $\mathcal{L}$ and N_d is evident in the LES and ModelE3 SCM results (Fig. A2). The match to the LES results is not a surprise for the default tuning, as such a comparison was considered in the ModelE3 model development phase. However, the decent match of Tun1, a product of the machine-learning tuning of the parent GCM, results partly from skill, in limiting the range of parameters that the machine learning explored, and partly from luck, as some parameter combinations devised by the machine learning did not result in such a good match to the LES result in these terms (not shown).

Code and data availability. Following acceptance, the analysis code and model output will be released with a code and data DOI

Author contributions. All authors contributed to the experiment design, model runs, data analysis, or manuscript writing.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Atmospheric Chemistry and Physics.

Acknowledgements. Bjorn Stevens's suggestion to use budget equations to characterize models' entrainment behavior provided the spark for this work. We thank Christopher Bretherton, Yao-Sheng Chen, Leo Donner, Graham Feingold, Tom Goren, Ed Gryspeerdt, Peter Kalmus, Jan Kazil, Adrian Lock, Roger Marchand, Daniel McCoy, Isabel McCoy, Roberto Mechoso, Brian Medeiros, Juan-Pedro Mellado, Yi Ming, Prasanth Prabhakaran, Phil Rasch, Christina Sackmann, Yunpeng Shan, Philip Stier, Shuaiqi Tang, João Teixeira, Chris Terai, Yoko Tsushima, Hui Wan, Rob Wood, Heng Xiao, Tak Yamaguchi, Mark Zelinka, Jianhao Zhang, Xiaoli Zhou, and four reviewers for comments and discussion. This work arises from the 2020 U.S. Climate Modeling Summit held virtually and chaired by Gavin Schmidt and Susanne Bauer. JM was supported by Office of Science, U.S. Department of Energy (DOE) Biological and Environmental Research as part of the Earth System Model Development (ESMD) program area and used resources of the National Energy Research Scientific Computing Center (NERSC), a

U.S. DOE Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under contract DE-AC02-05CH11231. The entrainment diagnostics were developed under the EAGLES project funded by the U.S. DOE ESMD program area. ASA, AMF, FT and SEB were supported by the NASA Modeling, Analysis, and Prediction Program and their computational resources were provided by the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. The Pacific Northwest National Laboratory (PNNL) is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RLO1830.

References

Ackerman, A., Kirkpatrick, M., Stevens, D., and Toon, O.: The impact of humidity above stratiform clouds on indirect aerosol climate forcing, Nature, 432, 1014–1017, https://doi.org/10.1038/nature03174, 2004.

Ackerman, A. S., vanZanten, M. C., Stevens, B., Savic-Jovcic, V., Bretherton, C. S., Chlond, A., Golaz, J.-C., Jiang, H., Khairoutdinov, M., Krueger, S. K., Lewellen, D. C., Lock, A., Moeng, C.-H., Nakamura, K., Petters, M. D., Snider, J. R., Weinbrecht, S., and Zulauf, M.: Large-Eddy Simulations of a Drizzling, Stratocumulus-Topped Marine Boundary Layer, Mon. Weather Rev., 137, 1083–1110, https://doi.org/10.1175/2008MWR2582.1, 2009.

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., McCoy, D. T., Myhre, G., Mülmenstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S. E., Sourdeval, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding Global Aerosol Radiative Forcing of Climate Change, Rev. Geophys., 58, e2019RG000 660, https://doi.org/10.1029/2019RG000660, 2020.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, JOURNAL OF HYDROLOGY, 249, 11–29, https://doi.org/10.1016/S0022-1694(01)00421-8, 2001.

Bogenschutz, P. A., Tang, S., Caldwell, P. M., Xie, S., Lin, W., and Chen, Y.-S.: The E3SM version 1 single-column model, Geosci. Model Dev., 13, 4443–4458, https://doi.org/10.5194/gmd-13-4443-2020, 2020.

Bretherton, C. S. and Park, S.: A new bulk shallow-cumulus model and implications for penetrative entrainment feedback on updraft buoyancy, J. Atmos. Sci., 65, 2174–2193, https://doi.org/10.1175/2007JAS2242.1, 2008.

Bretherton, C. S., Blossey, P. N., and Uchida, J.: Cloud droplet sedimentation, entrainment efficiency, and subtropical stratocumulus albedo, Geophys. Res. Lett., 34, https://doi.org/10.1029/2006GL027648, 2007.

Caldwell, P., Bretherton, C. S., and Wood, R.: Mixed-layer budget analysis of the diurnal cycle of entrainment in Southeast Pacific stratocumulus, J. Atmos. Sci., 62, 3775–3791, https://doi.org/10.1175/JAS3561.1, 2005.

Cesana, G., Del Genio, A. D., Ackerman, A. S., Kelley, M., Elsaesser, G., Fridlind, A. M., Cheng, Y., and Yao, M.-S.: Evaluating models' response of tropical low clouds to SST forcings using CALIPSO observations, Atmos. Chem. Phys., 19, 2813–2832, https://doi.org/10.5194/acp-19-2813-2019, 2019.

Cesana, Gregory, V., Ackerman, A. S., Fridlind, A. M., Silber, I., and Kelley, M.: Snow Reconciles Observed and Simulated Phase Partitioning and Increases Cloud Feedback, Geophys. Res. Lett., 48, e2021GL094 876, https://doi.org/10.1029/2021GL094876, 2021.

Christensen, M. W., Jones, W. K., and Stier, P.: Aerosols enhance cloud lifetime and brightness along the stratus-to-cumulus transition, Proc. Nat. Acad. Sci. USA, 117, 17 591–17 598, https://doi.org/10.1073/pnas.1921231117, 2020.

Christensen, M. W., Ma, P.-L., Wu, P., Varble, A. C., Mülmenstädt, J., and Fast, J. D.: Evaluation of aerosol-cloud interactions in E3SM using a Lagrangian framework, Atmos. Chem. Phys., 23, 2789–2812, https://doi.org/10.5194/acp-23-2789-2023, 2023.

Covert, J. A., Mechem, D. B., and Zhang, Z.: Subgrid-scale horizontal and vertical variation of cloud water in stratocumulus clouds: a case study based on LES and comparisons with in situ observations, Atmos. Chem. Phys., 22, 1159–1174, https://doi.org/10.5194/acp-22-1159-2022, 2022.

Elsaesser, G. S., O'Dell, C. W., Lebsock, M. D., Bennartz, R., Greenwald, T. J., and Wentz, F. J.: The Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP), J. Climate, 30, 10 193–10 210, https://doi.org/10.1175/JCLI-D-16-0902.1, 2017.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Feingold, G., Goren, T., and Yamaguchi, T.: Quantifying albedo susceptibility biases in shallow clouds, Atmos. Chem. Phys., 22, 3303–3319, https://doi.org/10.5194/acp-22-3303-2022, 2022.

Fons, E., Runge, J., Neubauer, D., and Lohmann, U.: Stratocumulus adjustments to aerosol perturbations disentangled with a causal approach, npj Clim. Atmos. Sci., 6, 130, https://doi.org/10.1038/s41612-023-00452-w, 2023.

Geoffroy, O., Brenguier, J.-L., and Burnet, F.: Parametric representation of the cloud droplet spectra for LES warm bulk microphysical schemes, Atmos. Chem. Phys., 10, 4835–4848, https://doi.org/10.5194/acp-10-4835-2010, 2010.

Gettelman, A., Morrison, H., Santos, S., Bogenschutz, P., and Caldwell, P. M.: Advanced Two-Moment Bulk Microphysics for Global Models. Part II: Global Model Solutions and Aerosol-Cloud Interactions, J. Climate, 28, 1288–1307, https://doi.org/10.1175/JCLI-D-14-00103.1, 2015.

Glassmeier, F., Hoffmann, F., Johnson, J. S., Yamaguchi, T., Carslaw, K. S., and Feingold, G.: An emulator approach to stratocumulus susceptibility, Atmos. Chem. Phys., 19, https://doi.org/10.5194/acp-19-10191-2019, 2019.

Golaz, J. C., Larson, V. E., and Cotton, W. R.: A PDF-based model for boundary layer clouds. Part I: Method and model description, J. Atmos. Sci., 59, 3540–3551, https://doi.org/10.1175/1520-0469(2002)059<3540:APBMFB>2.0.CO;2, 2002.

Golaz, J.-C., Salzmann, M., Donner, L. J., Horowitz, L. W., Ming, Y., and Zhao, M.: Sensitivity of the Aerosol Indirect Effect to Subgrid Variability in the Cloud Parameterization of the GFDL Atmosphere General Circulation Model AM3, J. Climate, 24, 3145–3160, https://doi.org/10.1175/2010JCLI3945.1, 2011.

Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., Bradley, A. M., Tang, Q., Maltrud, M. E., Forsyth, R. M., Zhang, C., Zhou, T., Zhang, K., Zender, C. S., Wu, M., Wang, H., Turner, A. K., Singh, B., Richter, J. H., Qin, Y., Petersen, M. R., Mametjanov, A., Ma, P.-L., Larson, V. E., Krishna, J., Keen, N. D., Jeffery, N., Hunke, E. C., Hannah, W. M., Guba, O., Griffin, B. M., Feng, Y., Engwirda, D., Di Vittorio, A. V., Dang, C., Conlon, L. M., Chen, C.-C.-J., Brunke, M. A., Bisht, G., Benedict, J. J., Asay-Davis, X. S., Zhang, Y., Zhang, M., Zeng, X., Xie, S., Wolfram, P. J., Vo, T., Veneziani, M., Tesfa, T. K., Sreepathi, S., Salinger, A. G., Eyre, J. E. J. R., Prather, M. J., Mahajan, S., Li, Q., Jones, P. W., Jacob, R. L., Huebler, G. W., Huang, X., Hillman, B. R., Harrop, B. E., Foucar, J. G., Fang, Y., Comeau, D. S., Caldwell, P. M., Bartoletti, T., Balaguru, K., Taylor, M. A., McCoy, R. B., Leung, L. R., and Bader, D. C.: The DOE E3SM Model Version 2: Overview of the Physical Model and Initial Model Evaluation, J. Adv. Model. Earth Syst., 14, e2022MS003 156, https://doi.org/10.1029/2022MS003156, 2022.

Gryspeerdt, E., Mülmenstädt, J., Gettelman, A., Malavelle, F. F., Morrison, H., Neubauer, D., Partridge, D. G., Stier, P., Takemura, T., Wang, H., Wang, M., and Zhang, K.: Surprising similarities in model and observational aerosol radiative forcing estimates, Atmos. Chem. Phys., 20, 613–623, https://doi.org/10.5194/acp-20-613-2020, 2020.

Guo, H., Golaz, J.-C., and Donner, L. J.: Aerosol effects on stratocumulus water paths in a PDF-based parameterization, Geophys. Res. Lett., 38, L17 808, https://doi.org/10.1029/2011GL048611, 2011.

Gustafson Jr., W. I., Vogelmann, A. M., Li, Z., Cheng, X., Dumas, K. K., Endo, S., Johnson, K. L., Krishna, B., Toto, T., and Xiao, H.: The Large-Eddy Simulation (LES) Atmospheric Radiation Measurement (ARM) Symbiotic Simulation and Observation (LASSO) Activity for Continental Shallow Convection, Bull. Amer. Meteorol. Soc., 101, E462–E479, https://doi.org/10.1175/BAMS-D-19-0065.1, 2020.

Hoffmann, F., Glassmeier, F., Yamaguchi, T., and Feingold, G.: Liquid Water Path Steady States in Stratocumulus: Insights from Process-Level Emulation and Mixed-Layer Theory, J. Atmos. Sci., 77, 2203–2215, https://doi.org/10.1175/JAS-D-19-0241.1, 2020.

Kalmus, P., Lebsock, M., and Teixeira, J.: Observational Boundary Layer Energy and Water Budgets of the Stratocumulus-to-Cumulus Transition, J. Climate, 27, 9155–9170, https://doi.org/10.1175/JCLI-D-14-00242.1, 2014.

Karset, I. H. H., Gettelman, A., Storelvmo, T., Alterskjaer, K., and Berntsen, T. K.: Exploring Impacts of Size-Dependent Evaporation and Entrainment in a Global Model, J. Geophys. Res., 125, e2019JD031 817, https://doi.org/10.1029/2019JD031817, 2020.

Kazil, J., Christensen, M. W., Abel, S. J., Yamaguchi, T., and Feingold, G.: Realism of Lagrangian Large Eddy Simulations Driven by Renalysis Meteorology: Tracking a Pocket of Open Cells Under a Biomass Burning Aerosol Layer, J. Adv. Model. Earth Syst., 13, e2021MS002 664, https://doi.org/10.1029/2021MS002664, 2021.

Larson, V. E.: CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere, https://doi.org/10.48550/arXiv.1711.03675, 2022.

Lee, L. A., Reddington, C. L., and Carslaw, K. S.: On the relationship between aerosol model uncertainty and radiative forcing uncertainty, Proc. Nat. Acad. Sci. USA, 113, 5820–5827, https://doi.org/10.1073/pnas.1507050113, 2016.

Lilly, D.: MODELS OF CLOUD-TOPPED MIXED LAYERS UNDER A STRONG INVERSION, Quart. J. Roy. Meteorol. Soc., 94, 292–&, https://doi.org/10.1002/qj.49709440106, 1968.

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

Ma, P.-L., Harrop, B. E., Larson, V. E., Neale, R. B., Gettelman, A., Morrison, H., Wang, H., Zhang, K., Klein, S. A., Zelinka, M. D., Zhang, Y., Qian, Y., Yoon, J.-H., Jones, C. R., Huang, M., Tai, S.-L., Singh, B., Bogenschutz, P. A., Zheng, X., Lin, W., Quaas, J., Chepfer, H., Brunke, M. A., Zeng, X., Mülmenstädt, J., Hagos, S., Zhang, Z., Song, H., Liu, X., Pritchard, M. S., Wan, H., Wang, J., Tang, Q., Caldwell, P. M., Fan, J., Berg, L. K., Fast, J. D., Taylor, M. A., Golaz, J.-C., Xie, S., Rasch, P. J., and Leung, L. R.: Better calibration of cloud parameterizations and subgrid effects increases the fidelity of the E3SM Atmosphere Model version 1, Geosci. Model Dev., 15, 2881–2916, https://doi.org/10.5194/gmd-15-2881-2022, 2022.

Mahfouz, N. et al.: Pre-industrial cloud albedo unconstrained by present-day correlations in E3SM v2, Atmos. Chem. Phys. Discuss., submitted.

McCoy, D. T., Field, P., Gordon, H., Elsaesser, G. S., and Grosvenor, D. P.: Untangling causality in midlatitude aerosol-cloud adjustments, Atmos. Chem. Phys., 20, 4085–4103, https://doi.org/10.5194/acp-20-4085-2020, 2020.

Medeiros, B. and Stevens, B.: Revealing differences in GCM representations of low clouds, Clim. Dynam., 36, 385–399, https://doi.org/10.1007/s00382-009-0694-5, 2011.

Mellado, J. P.: Cloud-Top Entrainment in Stratocumulus Clouds, Annual Review of Fluid Mechanics, 49, 145–169, https://doi.org/10.1146/annurev-fluid-010816-060231, 2017.

Mellado, J. P., Bretherton, C. S., Stevens, B., and Wyant, M. C.: DNS and LES for Simulating Stratocumulus: Better Together, J. Adv. Model. Earth Syst., 10, 1421–1438, https://doi.org/10.1029/2018MS001312, 2018.

Merk, D., Deneke, H., Pospichal, B., and Seifert, P.: Investigation of the adiabatic assumption for estimating cloud micro- and macrophysical properties from satellite and ground observations, Atmos. Chem. Phys., 16, 933–952, https://doi.org/10.5194/acp-16-933-2016, 2016.

Mülmenstädt, J. and Feingold, G.: The radiative forcing of aerosol–cloud interactions in liquid clouds: Wrestling and embracing uncertainty, Curr. Clim. Change Rep., 4, 23–40, https://doi.org/10.1007/s40641-018-0089-y, 2018.

Mülmenstädt, J. and Wilcox, L. J.: The Fall and Rise of the Global Climate Model, J. Adv. Model. Earth Syst., 13, e2021MS002 781, https://doi.org/10.1029/2021MS002781, 2021.

Mülmenstädt, J., Nam, C., Salzmann, M., Kretzschmar, J., L'Ecuyer, T. S., Lohmann, U., Ma, P.-L., Myhre, G., Neubauer, D., Stier, P., Suzuki, K., Wang, M., and Quaas, J.: Reducing the aerosol forcing uncertainty using observational constraints on warm rain processes, Science Adv., 6, eaaz6433, https://doi.org/10.1126/sciadv.aaz6433, 2020.

Mülmenstädt, J., Salzmann, M., Kay, J. E., Zelinka, M. D., Ma, P.-L., Nam, C., Kretzschmar, J., Hörnig, S., and Quaas, J.: An underestimated negative cloud feedback from cloud lifetime changes, Nat. Clim. Chang., 11, 508–513, https://doi.org/10.1038/s41558-021-01038-1, 2021.

Mülmenstädt, J. et al.: General circulation models simulate negative liquid water path–droplet number correlations, but anthropogenic aerosols still increase simulated liquid water path, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/egusphere-2024-4, 2024.

Proske, U., Ferrachat, S., Klampt, S., Abeling, M., and Lohmann, U.: Addressing Complexity in Global Aerosol Climate Model Cloud Microphysics, J. Adv. Model. Earth Syst., 15, e2022MS003 571, https://doi.org/10.1029/2022MS003571, 2023.

Randall, D.: Stratocumulus Cloud Deepening Through Entrainment, TELLUS SERIES A-DYNAMIC METEOROLOGY AND OCEANOGRAPHY, 36, 446–457, https://doi.org/10.1111/j.1600-0870.1984.tb00261.x, 1984.

Randall, D., Abeles, J., and Corsetti, T.: Seasonal simulations of the planetary boundary-layer and boundary-layer stratocumulus clouds with a general-circulation model, J. Atmos. Sci., 42, 641–676, https://doi.org/10.1175/1520-0469(1985)042<0641:SSOTPB>2.0.CO;2, 1985.

Regayre, L. A., Johnson, J. S., Yoshioka, M., Pringle, K. J., Sexton, D. M. H., Booth, B. B. B., Lee, L. A., Bellouin, N., and Carslaw, K. S.: Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF, Atmos. Chem. Phys., 18, 9975–10 006, https://doi.org/10.5194/acp-18-9975-2018, 2018.

Salzmann, M., Ming, Y., Golaz, J.-C., Ginoux, P. A., Morrison, H., Gettelman, A., Kraemer, M., and Donner, L. J.: Two-moment bulk stratiform cloud microphysics in the GFDL AM3 GCM: description, evaluation, and sensitivity tests, Atmos. Chem. Phys., 10, 8037–8064, https://doi.org/10.5194/acp-10-8037-2010, 2010.

Stephens, G.: Radiation profiles in extended water clouds .1. theory, J. Atmos. Sci., 35, 2111–2122, https://doi.org/10.1175/1520-0469(1978)035<2111:RPIEWC>2.0.CO;2, 1978.

Stevens, B.: Entrainment in stratocumulus-topped mixed layers, Quart. J. Roy. Meteorol. Soc., 128, 2663–2690, https://doi.org/10.1256/qj.01.202, 2002.

Stevens, B.: Atmospheric moist convection, Annual Review of Earth and Planetary Sciences, 33, 605–643, https://doi.org/10.1146/annurev.earth.33.092203.122658, 2005.

Stevens, B. and Feingold, G.: Untangling aerosol effects on clouds and precipitation in a buffered system, Nature, 461, 607–613, https://doi.org/10.1038/nature08281, 2009.

Stevens, B., Lenschow, D. H., Vali, G., Gerber, H., Bandy, A., Blomquist, B., Brenguier, J. L., Bretherton, C. S., Burnet, F., Campos, T., Chai, S., Faloona, I., Friesen, D., Haimov, S., Laursen, K., Lilly, D. K., Loehrer, S. M., Malinowski, S. P., Morley, B., Petters, M. D., Rogers, D. C., Russell, L., Savic-Jovac, V., Snider, J. R., Straub, D., Szumowski, M. J., Takagi, H., Thornton, D. C., Tschudi, M., Twohy, C., Wetzel, M., and van Zanten, M. C.: Dynamics and chemistry of marine stratocumulus - Dycoms-II, Bull. Amer. Meteorol. Soc., 84, 579–593, https://doi.org/10.1175/BAMS-84-5-579, 2003.

Terai, C. R., Pritchard, M. S., Blossey, P., and Bretherton, C. S.: The Impact of Resolving Subkilometer Processes on Aerosol-Cloud Interactions of Low-Level Clouds in Global Model Simulations, J. Adv. Model. Earth Syst., 12, e2020MS002 274, https://doi.org/10.1029/2020MS002274, 2020.

Tornow, F., Ackerman, A. S., and Fridlind, A. M.: Preconditioning of overcast-to-broken cloud transitions by riming in marine cold air outbreaks, Atmos. Chem. Phys., 21, 12 049–12 067, https://doi.org/10.5194/acp-21-12049-2021, 2021.

vanZanten, M. C., Stevens, B., Vali, G., and Lenschow, D. H.: Observations of drizzle in nocturnal marine stratocumulus, J. Atmos. Sci., 62, 88–106, https://doi.org/10.1175/JAS-3355.1, 2005.

Varble, A. C., Ma, P.-L., Christensen, M. W., Muelmenstaedt, J., Tang, S., and Fast, J.: Evaluation of liquid cloud albedo susceptibility in E3SM using coupled eastern North Atlantic surface and satellite retrievals, Atmos. Chem. Phys., 23, 13 523–13 553, https://doi.org/10.5194/acp-23-13523-2023, 2023.

von Bertalanffy, L.: The Theory of Open Systems in Physics and Biology, Science, 111, 23–29, https://doi.org/10.1126/science.111.2872.23, 1950.

Wood, R.: Stratocumulus Clouds, Mon. Weather Rev., 140, 2373–2423, https://doi.org/10.1175/MWR-D-11-00121.1, 2012.

Wyant, M. C., Bretherton, C. S., Chlond, A., Griffin, B. M., Kitagawa, H., Lappen, C.-L., Larson, V. E., Lock, A., Park, S., de Roode, S. R., Uchida, J., Zhao, M., and Ackerman, A. S.: A single-column model intercomparison of a heavily drizzling stratocumulus-topped boundary layer, J. Geophys. Res., 112, D24 204, https://doi.org/10.1029/2007JD008536, 2007.

Xie, S., Wang, Y.-C., Lin, W., Ma, H.-Y., Tang, Q., Tang, S., Zheng, X., Golaz, J.-C., Zhang, G. J., and Zhang, M.: Improved Diurnal Cycle of Precipitation in E3SM With a Revised Convective Triggering Function, J. Adv. Model. Earth Syst., 11, 2290–2310, https://doi.org/10.1029/2019MS001702, 2019.

Yamaguchi, T. and Randall, D. A.: Cooling of Entrained Parcels in a Large-Eddy Simulation, J. Atmos. Sci., 69, 1118–1136, https://doi.org/10.1175/JAS-D-11-080.1, 2012.

Zelinka, M. D., Klein, S. A., Qin, Y., and Myers, T. A.: Evaluating Climate Models' Cloud Feedbacks Against Expert Judgment, J. Geophys. Res., 127, e2021JD035 198, https://doi.org/10.1029/2021JD035198, 2022.

Zhang, G. and Mcfarlane, N.: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Center general-circulation model, Atmosphere-Ocean, 33, 407–446, https://doi.org/10.1080/07055900.1995.9649539, 1995.

Zhang, J. and Feingold, G.: Distinct regional meteorological influences on low-cloud albedo susceptibility over global marine stratocumulus regions, Atmos. Chem. Phys., 23, 1073–1090, https://doi.org/10.5194/acp-23-1073-2023, 2023.

Zhang, J., Zhou, X., Goren, T., and Feingold, G.: Albedo susceptibility of northeastern Pacific stratocumulus: the role of covarying meteorological conditions, Atmos. Chem. Phys., 22, 861–880, https://doi.org/10.5194/acp-22-861-2022, 2022.

Zhou, C. and Penner, J. E.: Why do general circulation models overestimate the aerosol cloud lifetime effect?: A case study comparing CAM5 and a CRM, Atmos. Chem. Phys., 17, 21–29, https://doi.org/10.5194/acp-17-21-2017, 2017.

Zhou, X. and Bretherton, C. S.: Simulation of Mesoscale Cellular Convection in Marine Stratocumulus: 2. Nondrizzling Conditions, J. Adv. Model. Earth Syst., 11, 3–18, https://doi.org/10.1029/2018MS001448, 2019.

Zhu, P., Bretherton, C. S., Köhler, M., Cheng, A. N., Chlond, A., Geng, Q. Z., Austin, P., Golaz, J. C., Lenderink, G., Lock, A., and Stevens, B.: Intercomparison and interpretation of single-column model simulations of a nocturnal stratocumulus-topped marine boundary layer, Mon. Weather Rev., 133, 2741–2758, https://doi.org/10.1175/MWR2997.1, 2005.

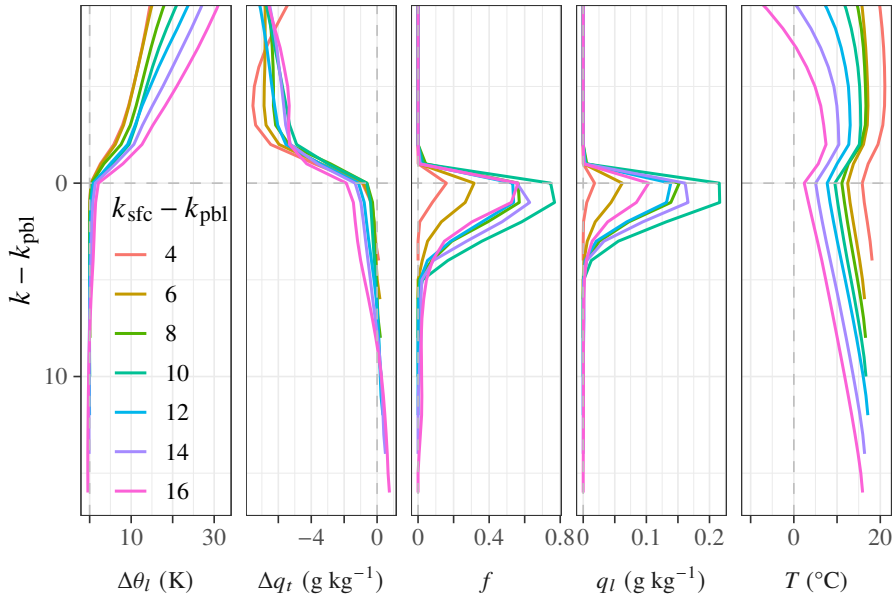

Figure 1. Composite vertical profiles in E3SM Sc conditions as defined in Sect. 2.2. To enable plotting of a cloud cover profile, the $f > 0.9$ requirement is not applied. Model-level differences are used as the vertical coordinate to avoid artificially smearing out the vertical gradients through remapping to a pressure or geometric height coordinate; model thickness varies with height and surface pressure but averages approximately 80 m in the profiles shown. The lowest level in the column at which the temperature increases with altitude is identified as the PBL top, with corresponding level number k_{pbl}. The vertical coordinate is model level referenced to PBL top, $k - k_{\mathrm{pbl}}$ (positive downward). Profiles are stratified by PBL depth, measured as the difference between the PBL-top model level and the lowermost model level k_{sfc}. (In E3SMv2, $k_{\mathrm{sfc}} = 72$.) Profiles of q_t and θ_l are shown as differences Δq_t and $\Delta\theta_l$ with respect to the mass-weighted vertical mean over the PBL.

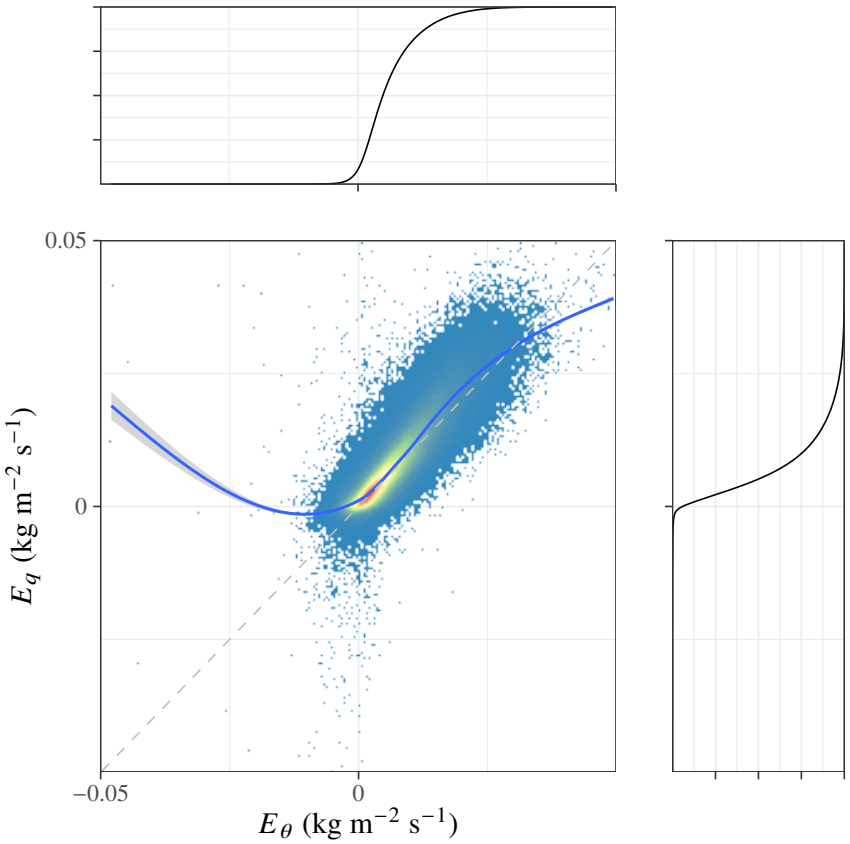

Figure 2. Physical consistency checks on E_q and E_θ. The central panel shows the joint probability density $P(E_q, E_\theta)$, along with a LOESS-smoothed mean E_q as a function of E_θ (blue line) and dashed gray 1:1 line. The outer panels show the marginal cumulative distributions of E_q and E_θ.

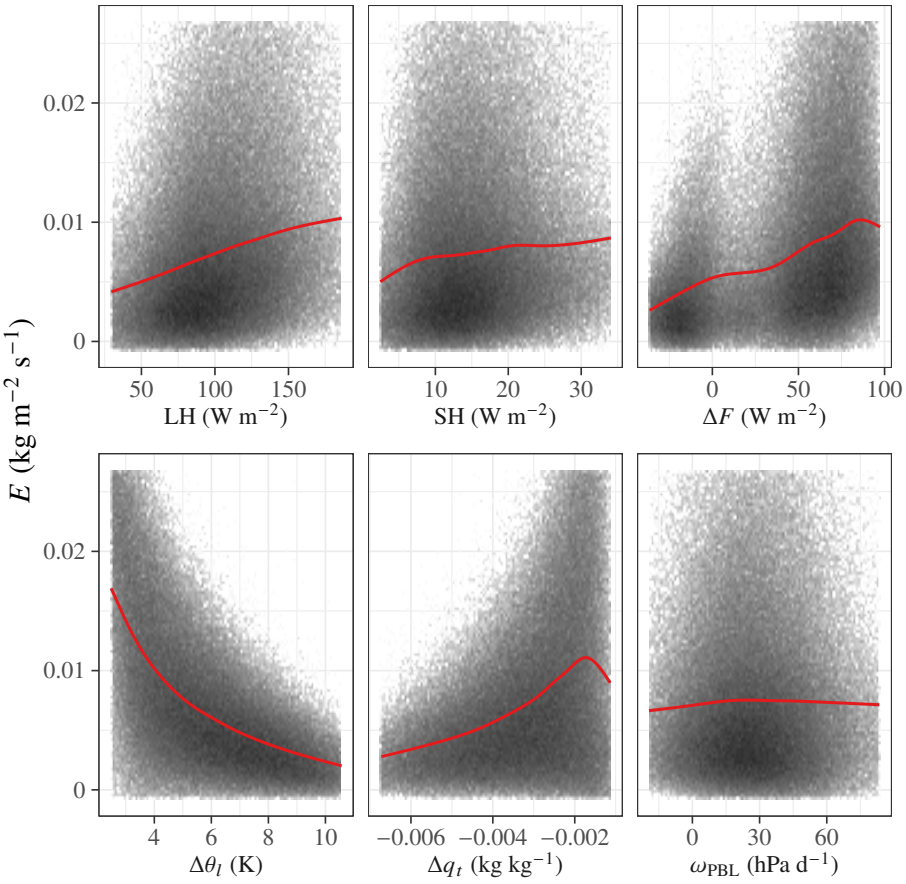

Figure 3. Dependence of entrainment on column properties: surface latent and sensible heat flux LH and SH, radiative cooling ΔF (cooling is positive), thermodynamic jumps across the inversion $\Delta\theta_l$ and Δq_t, and pressure vertical velocity at PBL top ω_{PBL}.

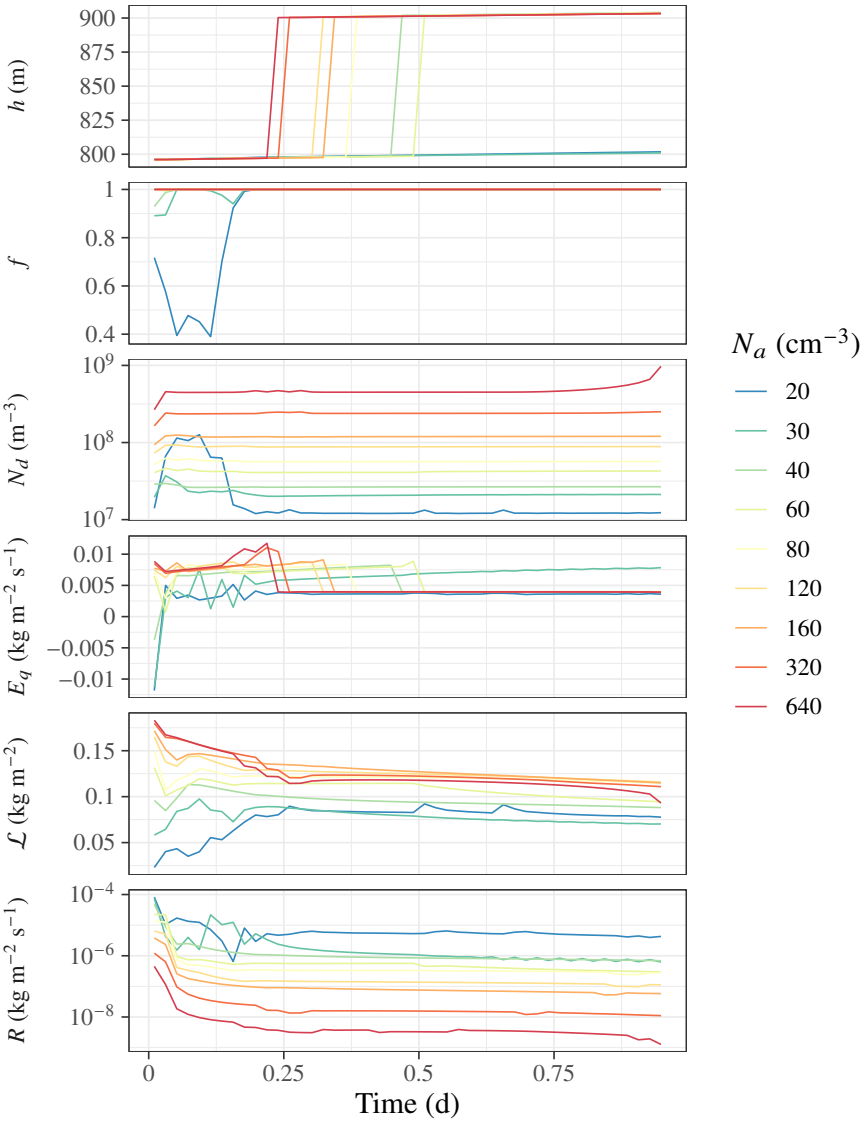

Figure 4. GISS ModelE3 SCM time series of PBL height (height of lowest model level with inverted temperature lapse) h, cloud cover f, mean droplet number N_d, entrainment E, liquid water path $\mathcal{L}$, and surface precipitation rate R for the DYCOMS-II RF02 experiment. Prescribed aerosol concentration is varied between $N_a = 20$ and 640 cm^{-3}.

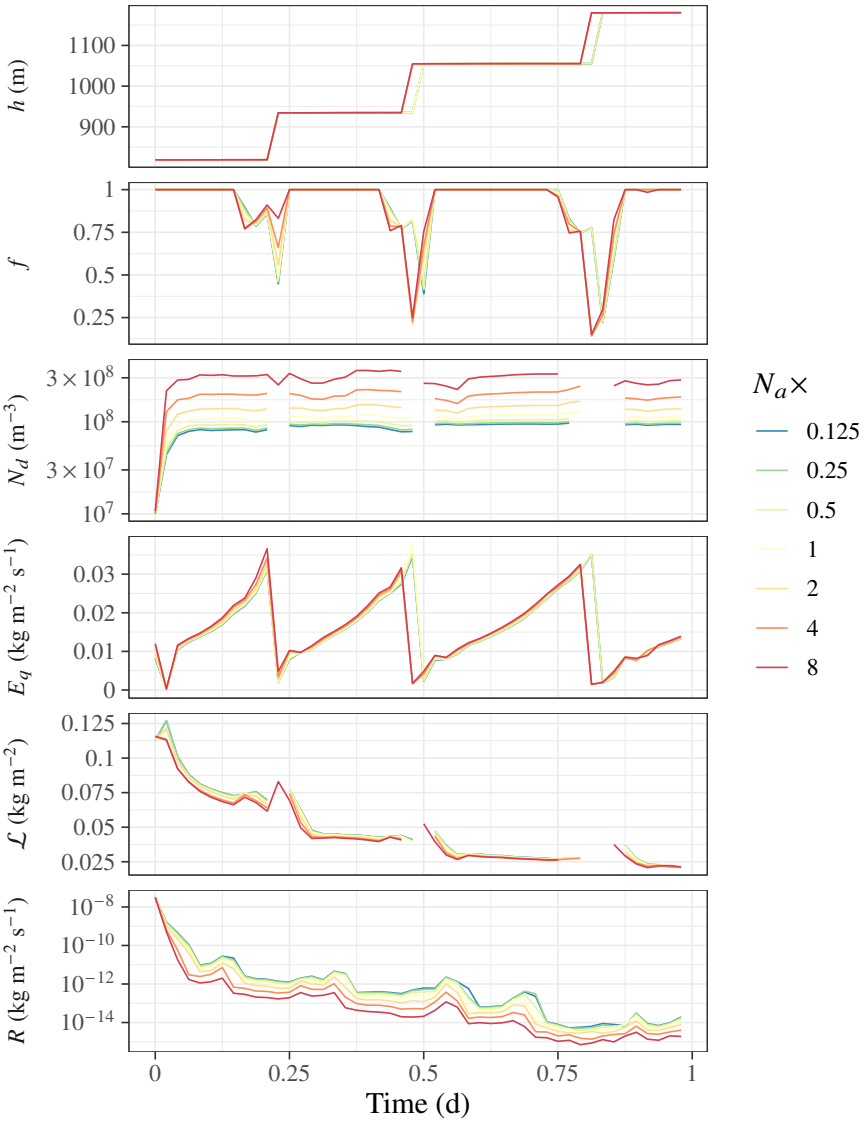

Figure 5. E3SM SCM time series for the DYCOMS-II RF02 setup. As in Fig. 4. Prescribed aerosol concentration is varied by a factor of 8 above and below its default value.

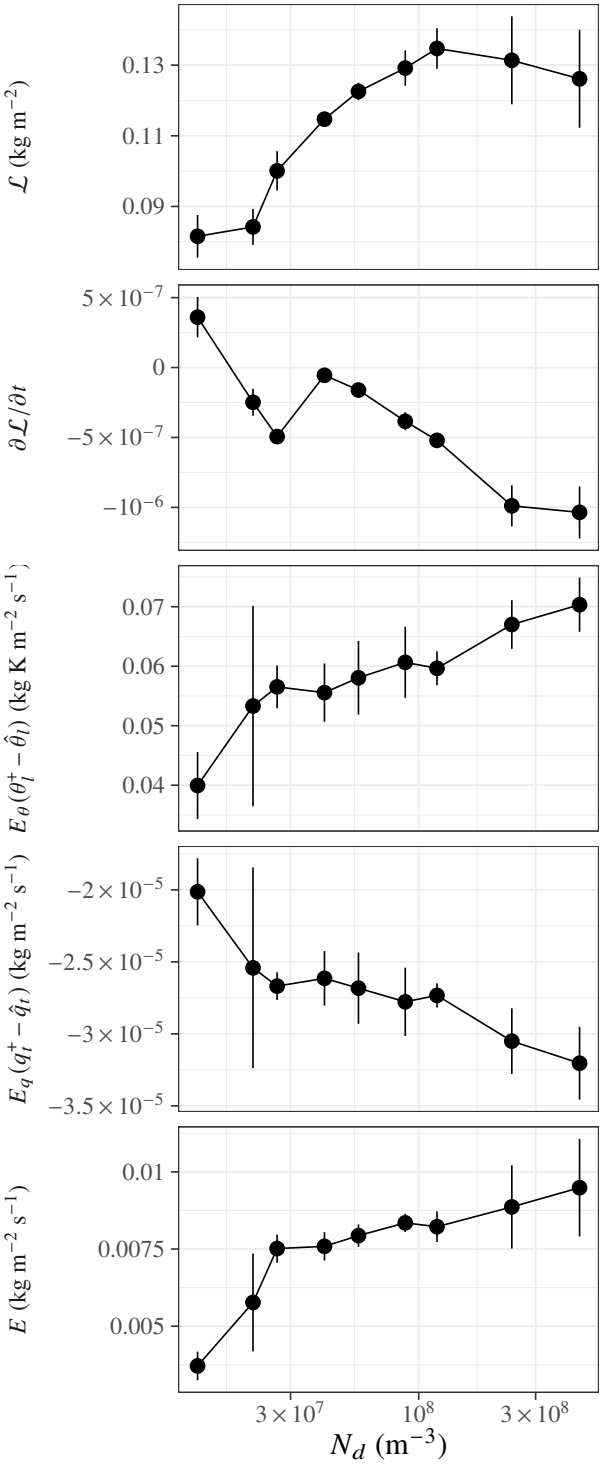

Figure 6. GISS SCM N_d relationships. State variables are averaged over the time period from 2–12 h; error bars indicate the standard deviation. The tendency $\partial\mathcal{L}/\partial t$ is calculated by linear regression over overcast conditions ($f > 0.9$) between the end of spinup (2 h) and 12 h; error bars indicate the standard error on the regression slope.

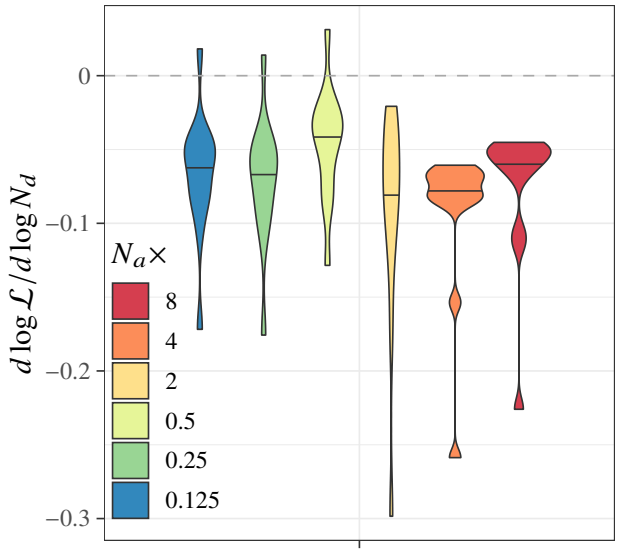

Figure 7. E3SM SCM $\partial\log\mathcal{L}/\partial\log N_d$. Susceptibilities are calculated at every time step from the $\delta\log\mathcal{L}$ and $\delta\log N_d$ in the perturbed-aerosol experiment relative to the default-aerosol experiment. The plot shows the density distribution of $\partial\log\mathcal{L}/\partial\log N_d$ over all time steps from the end of the spinup period (2 h) to 12 h, excluding periods when the cloud fraction drops below 0.9. Horizontal black lines across the density plots indicate the median.

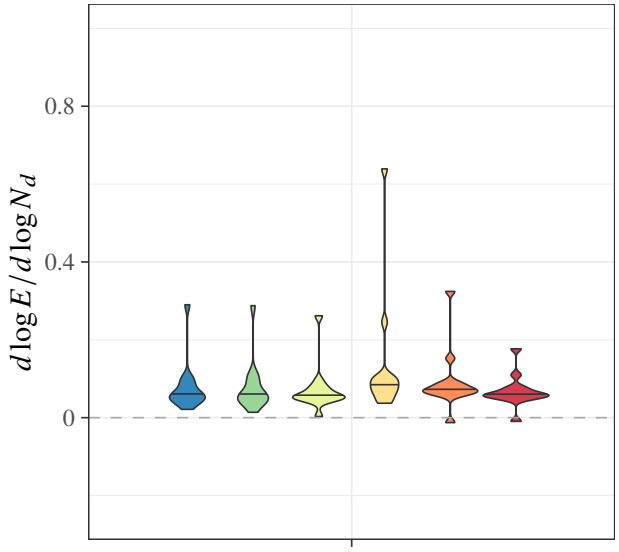

Figure 8. E3SM SCM $\partial \log E/\partial \log N_d$. As in Fig. 7 but showing the susceptibility of cloud-top entrainment.

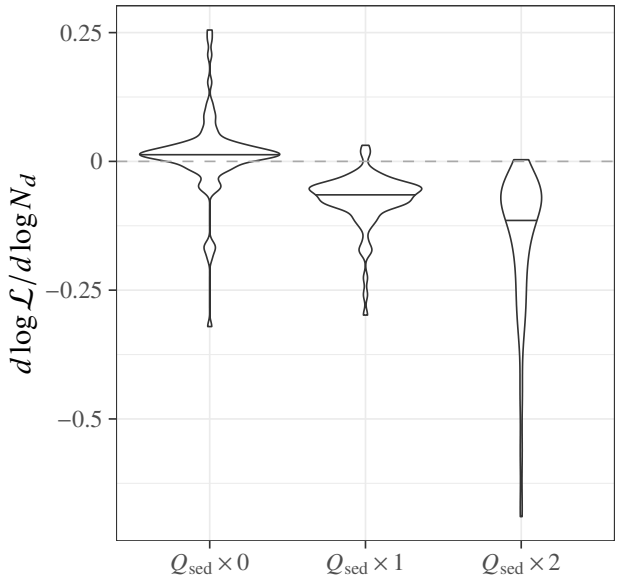

Figure 9. E3SM SCM $\partial \log \mathcal{L}/\partial \log N_d$ variation with size-dependent sedimentation. As in Fig. 7 but varying a scale factor applied to the sedimentation flux Q_{sed}.

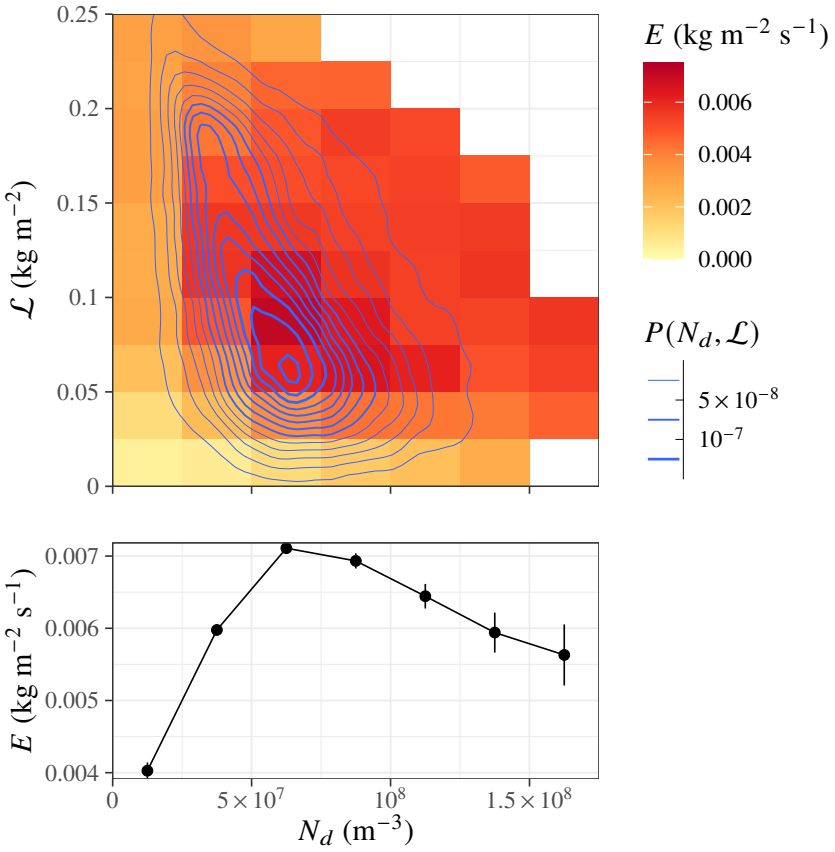

Figure 10. E3SM 3D atmosphere entrainment E. The top panel shows the dependence on $\mathcal{L}$ and N_d; only boxes with $n > 25$ points are included. Contours of the density $P(N_d, \mathcal{L})$ are overlaid. The bottom panel shows the dependence of E on N_d when the $\mathcal{L}$-dependent entrainment $E(N_d, \mathcal{L})$ is integrated over the $\mathcal{L}$ distribution; error bars indicate the standard error.

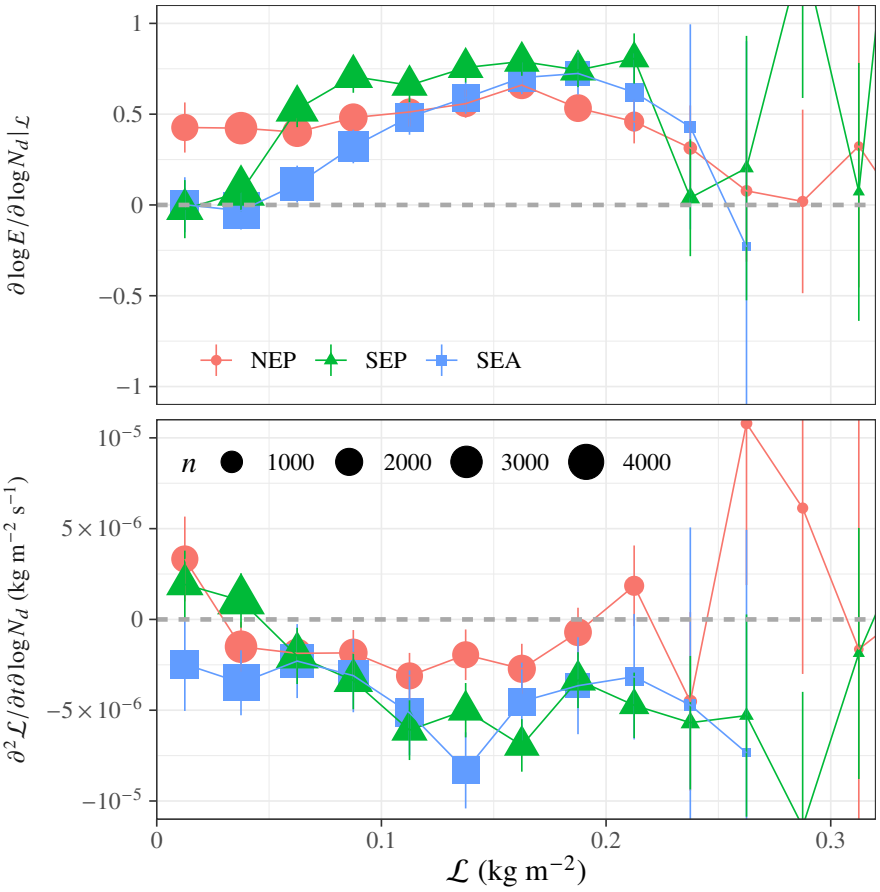

Figure 11. Susceptibility of entrainment and $\partial\mathcal{L}/\partial t$ to N_d. Susceptibility to N_d is calculated as linear regression slope of $\log E$ and $\partial\mathcal{L}/\partial t$ against $\log N_d$ over instantaneous PD statistics within each $\mathcal{L}$ bin.

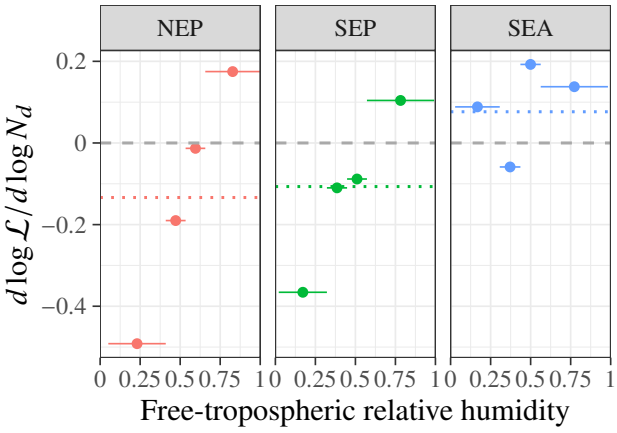

Figure 12. E3SM 3D atmosphere $d\log\mathcal{L}/d\log N_d$ stratified by free-tropospheric relative humidity quartiles. The susceptibility is calculated from the differences $\Delta\log\mathcal{L}$ and $\Delta\log N_d$ between PD and PI emissions runs averaged over each RH bin in each Sc region. Dashed lines indicate the regional Sc $d\log\mathcal{L}/d\log N_d$ mean integrated over free-tropospheric relative humidity.

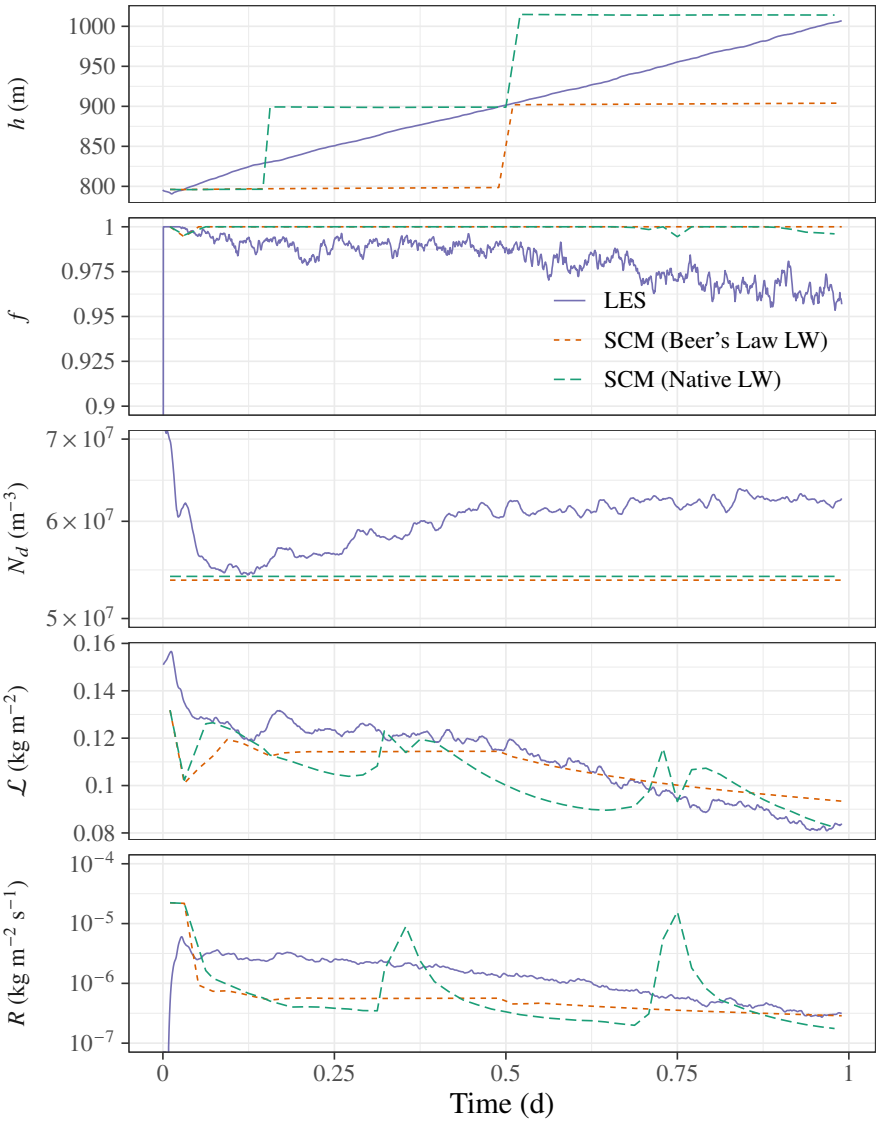

Figure A1. Evolution of domain-mean scalar diagnostics during 24-h simulations from DHARMA LES (solid blue line) and ModelE3 SCM, using Beer's Law parameterization of longwave raditive cooling per Ackerman et al. (2009) (red dotted line) and the ModelE3 native LW radiative transfer (green dashed line). The panels from the top depict domain-mean inversion height (location of maximum gradient in potential temperature below 5 km altitude), stratiform cloud cover (fraction of columns with opacity of at least 2.5 in the LES), cloud droplet concentration (average weighted by cloud water mixing ratio), liquid water path, and surface precipitation rate.

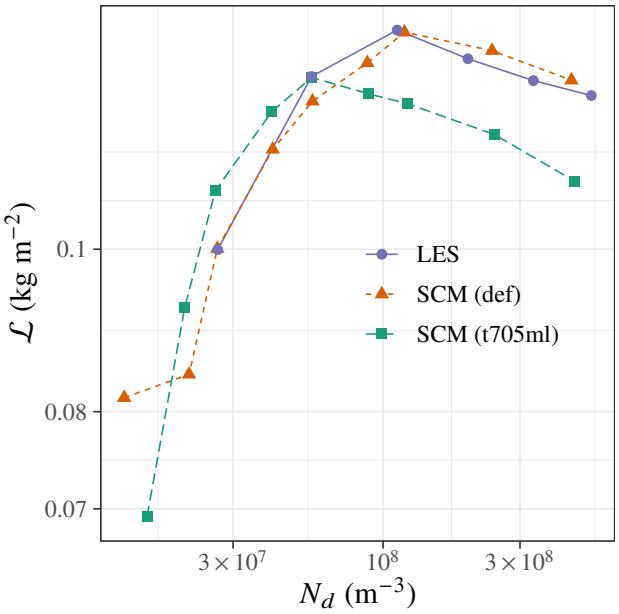

Figure A2. Domain mean $\mathcal{L}$ versus N_d (vertical average weighted by cloud water mixing ratio) for the DHARMA LES (solid blue line) and two ModelE3 SCM configurations: the default tuning (red dotted line) analyzed in this study and the machine-learning tuning Tun1 (dashed green) used in Part 1. The outputs are averaged over hours 2–12, and Beer's Law parameterization of longwave flux divergence is used for the LES and SCM.