# General circulation models simulate negative liquid water path–droplet number correlations, but anthropogenic aerosols still increase simulated liquid water path

Johannes Mülmenstädt[1], Edward Gryspeerdt[2], Sudhakar Dipu[3], Johannes Quaas[3], Andrew S. Ackerman[4], Ann M. Fridlind[4], Florian Tornow[4,5], Susanne E. Bauer[4], Andrew Gettelman[1], Yi Ming[6], Youtong Zheng[7,8], Po-Lun Ma[1], Hailong Wang[1], Kai Zhang[1], Matthew W. Christensen[1], Adam C. Varble[1], L. Ruby Leung[1], Xiaohong Liu[9], David Neubauer[10], Daniel G. Partridge[11], Philip Stier[12], and Toshihiko Takemura[13]

[1]Atmospheric, Climate and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA
[2]Grantham Institute - Climate Change and the Environment, Imperial College London, UK
[3]Leipzig Institute for Meteorology, Leipzig University, Leipzig, Germany
[4]NASA Goddard Institute for Space Studies, New York, NY, USA
[5]Columbia University Center for Climate System Research, New York, NY, USA
[6]Boston College, Boston, MA, USA
[7]Atmospheric and Oceanic Science Program, Princeton University, Princeton, NJ, USA
[8]Department of Earth and Atmospheric Science, University of Houston, Houston, TX, USA
[9]Texas A&M University, College Station, TX, USA
[10]Institute for Atmospheric and Climate Science, ETH Zürich, Zurich, Switzerland
[11]Department of Mathematics and Statistics, University of Exeter, UK
[12]Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, UK
[13]Research Institute for Applied Mechanics, Kyushu University, Fukuoka, Japan

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

**Abstract.** General circulation models' (GCMs) estimates of the liquid water path adjustment to anthropogenic aerosol emissions differ in sign from other lines of evidence. This reduces confidence in estimates of the effective radiative forcing of the climate by aerosol–cloud interactions (ERFaci). The discrepancy is thought to stem in part from GCMs' inability to represent the turbulence–microphysics interactions in cloud-top entrainment, a mechanism that leads to a reduction in liquid water in response to an anthropogenic increase in aerosols. In the real atmosphere, enhanced cloud-top entrainment is thought to be the dominant adjustment mechanism for liquid water path, weakening the overall ERFaci. We show that the latest generation of GCMs includes models that produce a negative correlation between present-day cloud droplet number and liquid water path, a key piece of observational evidence supporting liquid water path reduction by anthropogenic aerosols and one that earlier-generation GCMs could not reproduce. However, even in GCMs with this negative correlation, the increase in anthropogenic aerosols from preindustrial to present-day values still leads to an increase in simulated liquid water path due to the parameterized precipitation-suppression mechanism. This adds to the evidence that correlations in the present-day climate are not necessarily causal. We investigate sources of confounding to explain the noncausal correlation between liquid water path and droplet number. These results are a reminder that assessments of climate parameters based on multiple lines of evidence must carefully consider the complementary strengths of different lines when the lines disagree.

1

15

## 1  Introduction

Aerosol–cloud interactions (ACI) remain the greatest source of uncertainty in our estimates of anthropogenic perturbations
to Earth's energy budget (Boucher et al., 2014; Forster et al., 2021). In liquid clouds, an anthropogenic aerosol perturba-
tion essentially instantaneously alters the number of cloud droplets ($N_d$), changing cloud reflectance and thus the shortwave
20  radiation absorbed by the climate system, which exerts a radiative forcing on climate ("radiative forcing by aerosol–cloud
interactions" or RFaci; Twomey, 1977; Boucher et al., 2014). While our knowledge of RFaci is uncertain (Quaas et al., 2020),
an even thornier issue is cloud adjustments to the $N_d$ perturbation, where multiple processes acting at different scales from
cloud droplet to planetary circulation (Stevens and Feingold, 2009) result in a multiscale dynamics prediction problem that is
impervious to any one "silver bullet" solution (Mülmenstädt and Feingold, 2018). Estimates of ACI adjustments are, therefore,
25  based on multiple, and often conflicting, lines of evidence (Boucher et al., 2014; Bellouin et al., 2020; Forster et al., 2021).
Those lines of evidence are, broadly, modeling at the cloud process scale ("large eddy simulation" or LES), global modeling,
and observations at different scales.

    In the following, we focus on stratocumulus (Sc) clouds, which play a large role in the energy budget due to their high
albedo and frequent occurrence. Our understanding of adjustments in Sc is that two effects compete: an anthropogenic increase
30  in $N_d$ suppresses precipitation (Albrecht, 1989), increasing cloud liquid water path ($\mathcal{L}$); but the $N_d$ increase also promotes
increasing turbulent entrainment of subsaturated air at cloud top (Ackerman et al., 2004; Bretherton et al., 2007), decreasing
$\mathcal{L}$. These mechanisms are regime dependent; precipitation suppression only plays a role in clouds that would have precipitated
in the absence of the aerosol perturbation, and the entrainment mechanisms depend strongly on the turbulence generation
mechanisms, for example cloud-top radiative cooling. The regime dependence of the underlying processes leads to "process
35  fingerprints" in $N_d$–$\mathcal{L}$ space in LES (Hoffmann et al., 2020) for the very limited set of boundary conditions where LES is
available. Similar bifurcation behavior appears in satellite observations, where mean $\mathcal{L}$ as a function of $N_d$ first increases
in precipitating clouds, next reaches a peak that roughly coincides with the transition to nonprecipitating clouds, and then
decreases again (Gryspeerdt et al., 2019). There is evidence that this "inverted v" relationship between $\mathcal{L}$ and $N_d$ overestimates
the strength of the causal effect of $N_d$ on $\mathcal{L}$ (Gryspeerdt et al., 2019; Arola et al., 2022; Fons et al., 2023), but qualitatively it is
40  consistent with process understanding from LES. Integrated over all meteorological boundary conditions, the overall satellite
correlation between $\mathcal{L}$ and $N_d$ is negative. The satellite inverted v, satellite observations of natural laboratories (Christensen
et al., 2022) where the origin of the perturbation is evident (Malavelle et al., 2017; Toll et al., 2019), and process-modeling
lines of evidence lead to the assessment that the adjustment of $\mathcal{L}$ to anthropogenic aerosol is a reduction of $\mathcal{L}$, that is, a positive
contribution to the effective radiative forcing by ACI (ERFaci; Bellouin et al., 2020; Forster et al., 2021).
45  Global climate models – which, currently, means general circulation models (GCMs) run at roughly 1° latitude–longitude
spatial resolution – tell a different story. They would project an increase, rather than a decrease, in $\mathcal{L}$ when aerosols are
increased from preindustrial (PI) to present-day (PD) concentrations (Gryspeerdt et al., 2020). The GCM line of evidence is

discounted in multiline assessments because it conflicts with the other lines and because those lines are assumed to provide more reliable information. This assumption rests on the representation of the relevant processes in GCMs. In these models,

50   precipitation is initiated by a microphysical parameterization with an explicit dependence on $N_d$ (or, largely equivalent, droplet size), so that the $\mathcal{L}$ increase by precipitation suppression is explicitly parameterized. Reduced $\mathcal{L}$ by enhanced evaporation, on the other hand, depends critically on meter-scale or smaller interactions between turbulence, radiation, and microphysics at the cloud edge. These interactions fall between several parameterizations and are therefore tricky to formulate in GCMs. (As a perverse consequence, this causes us to fret that GCMs may be structurally incapable of representing turbulent entrainment

55   scales, while we often mistakenly consider the many-orders-of-magnitude-smaller-scale precipitation processes a parametric problem; e.g., Mülmenstädt et al., 2020, 2021).

In this work, we show that some Coupled Model Intercomparison Project 6 (CMIP6) era GCMs, unlike earlier model generations, are capable of producing inverted v $N_d$–$\mathcal{L}$ relationships in agreement with global observations and LES. Based on these PD correlations and on the $N_d$ change between PI and PD (i.e., mimicking the information available to observations-

60   based ERFaci estimates), these models predict a reduction in $\mathcal{L}$, which is consistent with assessments that use multiple lines of evidence. However, the causal effect of anthropogenic $N_d$ changes on $\mathcal{L}$, as diagnosed by model experiments where all climatic boundary conditions apart from aerosols are held fixed, remains as in previous GCM generations: an anthropogenic $N_d$ increase leads to an increase in average $\mathcal{L}$, consistent with a dominant role for the precipitation suppression mechanism parameterized in the model microphysics.

## 2   Data and methods

65

We use an ensemble of GCMs to perform fixed-sea surface temperature model experiments with PD and PI emissions, archive instantaneous aerosol and cloud information with sufficient frequency (3 h) to resolve the diurnal cycle and with sufficient length (1–5 years with the large-scale winds nudged to PD meteorology) to draw statistically robust conclusions. The model ensembles used are the CMIP5-era AeroCom indirect effect experiment (AeroCom IND3) simulations on the one hand and

70   four newer-version models prepared for CMIP6 on the other. The AeroCom models are described in Zhang et al. (2016); Ghan et al. (2016). The CMIP6-era models are the U.S. Department of Energy Exascale Earth System Model (E3SM) version 1 Atmosphere Model (EAMv1; Rasch et al., 2019), the NASA Goddard Institute for Space Studies (GISS) ModelE3 (Cesana et al., 2019, 2021) configuration Tun1, the Geophysical Fluid Dynamics Laboratory (GFDL) Atmospheric Model AM4.0 (Zhao et al., 2018), and the Community Earth System Model version 2 Community Atmosphere Model version 6 (CESM2-CAM6;

75   Gettelman et al., 2019). The CMIP6-era models were run for one year for the baseline experiment. E3SM was further run for 5 years for additional experiments that needed more data to perform stratification by confounding variables (see Sect. 3.3). For E3SM, the Cloud Feedback Model Intercomparison Project ObservaTon Simulator Package (COSP) satellite simulator (Pincus et al., 2012; Swales et al., 2018) mimicking the MODerate Resolution Imaging Spectroradiometer (MODIS) cloud retrievals (Platnick et al., 2017) and a number of vertically resolved fields were archived over a limited area over the northeast Pacific

80   (NEP) Sc region for further analysis of confounders.

## 2.1 Cloud selection

From the model output, we select liquid clouds, defined by the absence of ice (ice water path $< 10^{-3}$ kg m$^{-2}$ and ice cloud cover = 0) in the column. To mimic passive satellite analyses, as well as to simplify the application of entrainment diagnostics in part 2 of the series (Mülmenstädt et al., in prep.), we require near-overcast (liquid cloud cover > 0.9) conditions. For these

85   liquid clouds, we calculate "in-cloud" cloud-top $N_d$ and $\mathcal{L}$ by dividing the grid-mean $N_d$ and $\mathcal{L}$ by the projected cloud cover. Only clouds over ocean are considered in this analysis. We refer to these clouds as "overcast clouds".

In addition to these globally occurring overcast clouds, we also study smaller cloud subsets defined by dynamical regime following Medeiros and Stevens (2011). In this classification, the stratocumulus regime is based on vertical velocity $\omega$ at 700 and 500 hPa ($\omega_{700} > 10$ hPa d$^{-1}$ and $\omega_{500} > 10$ hPa d$^{-1}$) and lower tropospheric stability (LTS), which we define here as the

90   difference in potential temperature $\theta$ between 1000 hPa and 700 hPa ($\theta_{700} - \theta_{1000} > 18.55$ K). We further restrict the clouds to occur in grid boxes where these conditions are met at least 30% of the time, which serves to select the subtropical Sc regions. The occurrence fraction $f_{\mathrm{Sc}}$ of these conditions is shown in Fig. 1. In addition to the Medeiros and Stevens (2011) requirements, all of the above-mentioned warm cloud criteria are applied. We refer to these clouds as "Sc regime clouds".

## 2.2 Analysis methods

95   From $\mathcal{L}$ and $N_d$, we construct the conditional probability $P(\mathcal{L}|N_d)$ following Gryspeerdt et al. (2019). For ease of comparison among models and configurations, we collapse the two-dimensional $P(\mathcal{L}|N_d)$ into one dimension by calculating the geometric-mean $\mathcal{L}$ in each $N_d$ bin, also following Gryspeerdt et al. (2019).

For the MODIS simulator analysis in Sect. 3.3.3, we transform the simulated $\tau$ and droplet effective radius ($r_e$) into $N_d$ and $\mathcal{L}$ using a power-law relationship for adiabatic updrafts with constant $N_d$ (Brenguier et al., 2000; Bennartz, 2007; Painemal

100   and Zuidema, 2011; Grosvenor et al., 2018):

$$N_d = \frac{\sqrt{5}}{2\pi k \sqrt{\rho_w Q}} \sqrt{f_{\mathrm{ad}} \Gamma} \tau^{1/2} r_e^{-5/2} \tag{1}$$

$$\mathcal{L} = \frac{5}{9} \rho_w \tau r_e, \tag{2}$$

where we take the ratio $k = (r_v/r_e)^3$ between volumetric mean radius $r_v$ cubed and effective radius cubed to be 1, subadiabatic factor $f_{\mathrm{ad}} = 1$, scattering efficiency $Q = 2$, and adiabatic condensation rate $\Gamma = 2 \times 10^{-6}$ kg m$^{-4}$. These assumptions minimize

105   the complications involved in showing results that are mostly power-law behavior independent of these constant factors. (This does neglect important modifications that can arise if these factors are not, in fact, constant; Varble et al., 2023).

To analyze confounding by planetary boundary layer (PBL) depth (Sect. 3.3.2), we identify the top of the Sc-like boundary layer by the first model level where temperature increases with height in Sc-regime overcast columns. This produces well-mixed profiles of liquid-water potential temperature $\theta_l$ and total water mixing ratio $q_w$. (Other definitions of PBL top, i.e.,

110   the model level of greatest gradient in $\theta_l$ or $q_w$, yield very similar results.) As we will see in Sect. 3.3.2, cloud and aerosol properties are remarkably stratified by PBL depth in E3SM; to keep the properties as distinct as possible as a function of

PBL depth, we retain the native model vertical discretization instead of converting the hybrid pressure levels to pressure or geometric height.

Table 1 summarizes the emissions, cloud selection, and model run duration for each experiment.

## 3  Results and discussion

In Fig. 2, we show the behavior of the AeroCom IND3 (CMIP5-era) GCMs in $N_d$–$\mathcal{L}$ space: with the exception of one model, $\mathcal{L}$ increases monotonically as a function of $N_d$. In some models, the slope decreases at high $N_d$, but only one model (HadGEM) has quantitatively similar behavior to the inverted v satellite $N_d$–$\mathcal{L}$ plot. The behavior of these models (with the exception of HadGEM) is consistent with the interpretation that the predominant mechanism linking $\mathcal{L}$ and $N_d$ is precipitation suppression.

### 3.1  CMIP6-era models produce inverted v $N_d$–$\mathcal{L}$ relationships

A funny thing happened on the way to CMIP6: three of the four US CMIP6-era GCMs have an inverted v with a pronounced negative slope. The behavior of these models is contrasted with the AeroCom models' behavior in Fig. 3. The geographic distribution of the regression slope between $\log \mathcal{L}$ and $\log N_d$ is predominantly negative in the models with an inverted v (Fig. 4), as Gryspeerdt et al. (2019) found in satellite retrievals.

One of these models (ModelE) was designed to better represent the entrainment behavior to which the negative slope is attributed in process-scale modeling. The other two (CAM6 and EAMv1), however, were not; if the negative slope is due to an entrainment ACI mechanism, it is an emergent behavior not explicitly parameterized into the turbulence scheme. It is doubly surprising that these models produce a negative slope considering that their closely related predecessor, CAM5.3-CLUBB-MG2, was part of the AeroCom ensemble and showed, at best, a slightly negative relationship between $N_d$ and $\mathcal{L}$.

### 3.2  The negative correlation between $N_d$ and $\mathcal{L}$ does not predict the sign of PI to PD change in $\mathcal{L}$

The bulk of the $N_d$ population lies in the part of the inverted v with a negative $N_d$–$\mathcal{L}$ correlation. If we regarded this relationship as indicative of a causal influence of $N_d$ on $\mathcal{L}$ – that is, that an increase in $N_d$ causes $\mathcal{L}$ to decrease – then we would predict a decrease in $\mathcal{L}$ as $N_d$ increases from its PI value to its PD value due to anthropogenic emissions.

We can compare the change in $\mathcal{L}$ predicted by the $N_d$–$\mathcal{L}$ correlation in PD internal variability to the outcome of a model experiment designed to measure the causal effect of $N_d$ on $\mathcal{L}$. This experiment fixes all climatic boundary conditions affecting cloud state (i.e., solar constant, greenhouse gases, and sea-surface temperature) with the exception of anthropogenic aerosols. The change in $\mathcal{L}$ in this experiment can therefore only be due to the anthropogenic aerosol emissions change. This model experiment shows that the causal effect of the $N_d$ increase is to increase $\mathcal{L}$ on average, contradicting the prediction of a decrease in $\mathcal{L}$ based on PD internal variability (Fig. 5). The correlation seen in PD internal variability in these models therefore cannot be causal. Plotting the correlations within PD and PI, as shown in Fig. 6, provides a glimpse at what is happening instead: a secular increase in $N_d$ does not lead to a secular reduction in $\mathcal{L}$ by shifting the $\mathcal{L}$ population along the correlation line, as would be expected for a causal relationship. Instead, the correlation line shifts along with the secular shifts in $N_d$ and

$\mathcal{L}$ ( mostly to the right given that the change in $N_d$ is far greater than the change in $\mathcal{L}$) in a way that is not predicted by the correlation line itself.

145     This contradiction raises three questions. First, what produces the noncausal negative $N_d$–$\mathcal{L}$ correlation? We provide a few hypotheses in the following section. Second, considering that these models can replicate the observed PD correlation, what can we infer about the causality of the relationship in observations, where we are unable to conduct direct experimental tests of causality? We discuss this question in Sect. 3.4. Third, is any part of the negative relationship between $N_d$ and $\mathcal{L}$ in the models causal? Any such causal mechanism would have to involve a direct or indirect $N_d$-dependence in cloud-top

150     entrainment. In ModelE, the Bretherton and Park (2009) turbulence scheme provides an explicit entrainment closure. Guo et al. (2011) have shown that the combination of the Cloud Layers Unified By Binormals (CLUBB; Larson and Golaz, 2005; Golaz et al., 2007) cloud and turbulence scheme and the Morrison–Gettelman microphysics scheme (Morrison and Gettelman, 2008; Salzmann et al., 2010) can reproduce entrainment-mediated enhanced evaporation at high $N_d$ in single-column experiments. This behavior has not been documented in three-dimensional GCM experiments, but CAM6 and EAMv1 use related cloud–

155     turbulence (Bogenschutz et al., 2013; Larson, 2022) and cloud–microphysics (Gettelman, 2015) schemes, so it is conceivable that $N_d$-dependent entrainment mechanisms contribute to the $N_d$–$\mathcal{L}$ relationship in these three models. A deeper investigation of this question merits a separate paper (part 2 of this series, Mülmenstädt et al., in prep.).

### 3.3   Sources of covariability that produce noncausal $N_d$–$\mathcal{L}$ relationships

Noncausal relationships between two variables often originate from a third (possibly unobserved) variable that exerts a causal
160     relationship on the two variables being correlated. This third variable is termed a "confounding variable" (Pearl and Mackenzie, 2018). In its most striking form, confounding can lead to a sign reversal between causation and correlation, for example in Simpson's paradox (Simpson, 1951; Feingold et al., 2022). Cloud properties respond strongly to the circulation at the scales of the Sc cellular organization (mesoscale) and greater. Thus, the meso- to synoptic-scale circulation is a natural place to look for confounding variables that lead to noncausal correlations between cloud properties.

165     #### 3.3.1   Mesoscale cloud regimes

Mesoscale circulation manifests as cloud "regimes" (e.g., Rossow et al., 2005; Gryspeerdt and Stier, 2012; Muhlbauer et al., 2014; Unglaub et al., 2020). ACI mechanisms likely differ between cloud regimes (e.g., Mülmenstädt and Feingold, 2018; Possner et al., 2020; Dipu et al., 2022). This could result in different $N_d$–$\mathcal{L}$ slopes in open- or closed-cell Sc or shallow cumulus or, as the positive- and negative-sloped legs of the inverted v relationship perhaps show, in precipitating and nonprecipitating

170     cloud regimes. Due to GCMs' coarse resolution, it is doubtful that they can correctly represent these mesoscale cloud regimes, their ACI mechanisms, or their coupling to the circulation. Nevertheless – or perhaps precisely because we can probably discount cloud-scale causal links between $N_d$ and $\mathcal{L}$ due to the mismatch with the GCM resolved scale – we can use GCMs to test whether the existence of cloud regimes is, on its own, a confounding mechanism for the $N_d$–$\mathcal{L}$ relationship.

        To assess whether regime-induced confounding effects may exist in the model $N_d$–$\mathcal{L}$ relationship, we stratify the E3SM
175     model clouds by surface rain rate. These bins of rain rate are our stand-in for precipitation regimes. We focus on the surface

rain rate because, unlike mesoscale morphological regime definitions (which are subgrid scale in the GCM), the precipitating–nonprecipitating regime delineation has a somewhat clear analog in the GCM. Because the model rain rate has a very long low tail, we do not attempt to define a binary nonprecipitating versus precipitating categorization but rather divide the cloud sample into quantiles of rain rate. Specifically, we use sextiles, balancing the need for a meaningful range of rain rates with the need to maintain a large sample of clouds within each bin. The CloudSat precipitation detection sensitivity at the GCM spatial resolution ($\approx 0.01$ mm d$^{-1}$; Stephens et al., 2010) falls roughly into the third rain rate bin, so, by this definition, half the bins approximately represent precipitating and half nonprecipitating clouds.

Figure 7 shows the results. The model, perhaps unrealistically, produces clouds that generate surface-reaching rain at all droplet concentrations; however, in bins with higher rain rates, the $N_d$ distribution is noticeably lower, as might be expected from the negative-exponent power law that parameterizes the autoconversion of cloud water to rain, and as is expected from observations (Pawlowska and Brenguier, 2003; Comstock et al., 2004). At the same time, $\mathcal{L}$ is higher in bins with higher rain rates, again as might be expected from the parameterized autoconversion and accretion. Superimposing the bin-mean $N_d$ and $\mathcal{L}$ for each rain-rate bin on the unbinned $N_d$–$\mathcal{L}$ distribution, we find that the negative correlation among the bin means echoes the unbinned correlation. This is the case even though, in very classic Simpson (1951) fashion, the correlations within five out of the six $R$ bins are positive. Thus, the opposing influences of $N_d$ and $\mathcal{L}$ on rain rate can, without any involvement of entrainment or evaporation mechanisms, generate a noncausal negative correlation between $N_d$ and $\mathcal{L}$.

We note that the mechanism generating this noncausal correlation is unusual. The contrast with the causal precipitation suppression mechanism is clear: there, causation runs from $N_d$ to autoconversion to $\mathcal{L}$. Here, causation runs jointly from both $N_d$ and $\mathcal{L}$ to precipitation. How or whether the causal chain then returns from precipitation to $N_d$ and $\mathcal{L}$, as in the classic confounding mechanism, is an open question.

We further note that precipitation already appears to have a qualitative effect on the model's $N_d$–$\mathcal{L}$ relationship at rain rates far below the CloudSat sensitivity threshold: even in the second-lowest $R$ bin, the correlation between $N_d$ and $\mathcal{L}$ is already positive. This suggests that the parameterized precipitation may exert such a strong influence on ACI even for clouds with low precipitation rate that other ACI adjustment mechanisms, while they may in principle be represented in the model, could be so overwhelmed by the parameterized precipitation suppression that their effect is not discernible in the climate response.

### 3.3.2 Synoptic-scale airmass advection

At the synoptic to planetary scales, covariability between cloud and aerosol properties can lead to spurious correlations in ACI metrics (Grandey and Stier, 2010). Synoptic-scale meteorological covariability can take the form of continental versus marine airmass advection. When an airmass originates over land, it typically has higher temperature, lower relative humidity (contributing to lower $\mathcal{L}$), and higher aerosol concentration (contributing to higher $N_d$) than when an airmass originates over ocean. This contrast between airmasses creates an anticorrelation between $N_d$ and $\mathcal{L}$ even in the absence of any causal effect of $N_d$ on $\mathcal{L}$ (Brenguier et al., 2003). Additionally, sea surface temperature is coldest and climatological subsidence strongest, near the coast, resulting in shallow marine boundary layers. The model's conception of this synoptic-scale covariability in space can be seen in Fig. 8, with shallow boundary layers and high cloud condensation nuclei (CCN) concentrations near shore and

210 deeper boundary layers with low CCN farther offshore. A similar covariability exists at particular locations in time. Figure 9 illustrates the mechanism in the NEP Sc region: at any given location, PBL depth and CCN concentration are strongly linked via the synoptic-scale circulation. Presumably the position of the anticyclonic subtropical subsidence governs both the PBL depth and whether continental or maritime air is advected.

To assess the synoptic meteorological confounding effect, we stratify the E3SM model clouds in the NEP Sc region by PBL

215 depth. We choose PBL depth as the confounding variable because it appears to act as a proxy for airmass "continentality" in the model (Fig. 8), without a direct parameterized relationship to either aerosols or cloud. PBL depth is nevertheless strongly correlated with both CCN concentration (temporal- and regional-mean vertical profiles are shown in Fig. 10) and $\mathcal{L}$. For a fairly wide range of PBL depths (representing the central 90% of the PBL depth distribution for Sc-regime cloud columns), the relationship between mean $N_d$ and mean $\mathcal{L}$ stratified by PBL depth mimics the slope of the unstratified $N_d$–$\mathcal{L}$ relationship

220 quite closely (Fig. 11). Based on this, it is plausible that synoptic-scale meteorological covariability contributes substantially to the overall negative $N_d$–$\mathcal{L}$ correlation in the model.

We note several caveats. The synoptic-scale covariability of aerosol advection and PBL depth is a feature of the general circulation and can therefore be expected to be modeled reasonably well in a GCM. However, the interaction of the advected aerosol with boundary-layer clouds depends on mixing between the free troposphere and the boundary layer, which is likely

225 much less well represented in GCMs that have coarse vertical resolution. Whether the synoptic-scale confounding signature in the model mimics the real atmosphere is therefore uncertain. Further, the synoptic-scale covariability differs depending on the geographic particulars of each Sc basin; we have only analyzed the NEP Sc in detail. Finally, while the PBL depth-stratified negative $N_d$–$\mathcal{L}$ relationship in the model is consistent with observational analyses (e.g., Fons et al., 2023), the model does not reproduce the weakening of the $N_d$–$\mathcal{L}$ correlation within each PBL depth bin (not shown) that is found in observations

230 (Possner et al., 2020; Fons et al., 2023).

### 3.3.3 Phase-space boundaries

Correlations between $N_d$ and $\mathcal{L}$ can also arise simply because not all parts of the $N_d$–$\mathcal{L}$ phase space are equally accessible to clouds. This can be illustrated by applying the MODIS simulator (Pincus et al., 2012) to the model. The MODIS simulator provides optical thickness $\tau$ and droplet effective radius $r_e$ diagnosed consistently with the MODIS satellite cloud retrievals

235 (Platnick et al., 2017). Power-law adiabatic relationships $\mathcal{L}(r_e, \tau)$ and $N_d(r_e, \tau)$ can be used to transform the MODIS output into $N_d$–$\mathcal{L}$ space (e.g., Dipu et al., 2022). In logarithmic coordinates, this is a linear transformation, yielding the correlation shown in Fig. 12. This S-curve correlation, like the model-native $N_d$–$\mathcal{L}$ correlation, shows a steep rise in $\mathcal{L}$ at low $N_d$ and a steep drop at moderate $N_d$. It also shows another steep rise at high $N_d$ that the model may hint at but does not exhibit clearly. Investigating the data before and after the coordinate transformation to $N_d$–$\mathcal{L}$ space is instructive. In $\log \tau$–$\log r_e$ space, the

240 MODIS simulator output falls within a rectangle, bounded by the limits the model prescribes on its clouds. Upon transformation to $\log \mathcal{L}$–$\log N_d$ space, the population is bounded by a parallelogram (see the isolines in Fig. 12). These limits on the phase space strongly sculpt the behavior of the mean $\log \mathcal{L}$ as a function of $\log N_d$, because the parts of phase space that are not populated do not contribute to the mean $\mathcal{L}$ as a function of $N_d$.

### 3.4 Persistent disagreement with other lines of evidence

245   Before these results, it was only logical to discount the GCM evidence on the basis that it could not reproduce the observed the $N_d - \mathcal{L}$ relationship in PD internal variability. Now that some GCMs match the other lines of evidence in PD internal variability, what do we make of the fact that the disagreement on the sign of the causal climatic $\mathcal{L}$ adjustment to RFaci persists?

    In observations, it is more difficult to establish causality than in the GCMs, where it is as simple as changing the aerosol emissions while fixing all other boundary conditions. The most reliable causal evidence in observations comes from observational

250  natural laboratories where the aerosol perturbation is known and an unperturbed control can be identified clearly (Christensen et al., 2022). Such laboratories indicate unchanged or reduced $\mathcal{L}$ in the perturbed clouds (Malavelle et al., 2017; Toll et al., 2019; Diamond et al., 2020). But such laboratories are rare, and there is no rigorous extrapolation from these laboratories to the full diversity of cloud regimes found in the climate. The most representative observations, that is, the global satellite-retrieved inverted v correlations, have the opposite problem: they are representative, but are the correlations causal? The correlation is

255  more negative than the estimate of causal interannual $\mathcal{L}$ response to $N_d$ perturbations using an effusive volcano as a laboratory (Gryspeerdt et al., 2019, albeit for shallow Cu rather than Sc). Arola et al. (2022) argue that satellite $N_d - \mathcal{L}$ correlations are negatively biased not only by covariability confounding but also by retrieval errors. Fons et al. (2023) applied a causal network approach to the temporal evolution in geostationary satellite data and found that the causal negative $N_d - \mathcal{L}$ relationship is weaker than the $N_d - \mathcal{L}$ correlation. Strong regional increasing and declining trends on multidecadal timescales in the satellite

260  record may also contribute to disentangling covariability and causality (Quaas et al., 2022).

    In LES, as in GCMs, causality can be established by varying aerosol concentration while keeping the other boundary conditions constant. This provides very clear evidence that precipitation suppression and entrainment feedbacks lead to process fingerprints of positive and negative $\mathcal{L}$ tendencies in $N_d - \mathcal{L}$ space (Hoffmann et al., 2020) that translate into steady Sc states (Glassmeier et al., 2021). But these LES experiments are expensive, so boundary conditions are carefully curated to a very small

265  subset of the high-dimensional space of meteorological conditions present in the climate. We simply do not know whether the process fingerprints would be as unambiguous if a broader spectrum of boundary conditions were simulated or if the clouds were able to interact with larger scales in the multiscale climate problem (Kazil et al., 2021) instead of evolving to a steady state.

    In summary, GCMs are still the odd ones out in their negative $\mathcal{L}$ adjustment component of ERFaci. The observational and

270  LES modeling lines of evidence have clear confounding and representativeness problems. Are these problems severe enough to flip the sign of the adjustment? It seems unlikely, but our GCM results show that it is possible; addressing the representativeness and confounder questions in the other lines of evidence thus takes on a renewed urgency.

### 4 Conclusions

  Mülmenstädt and Wilcox (2021) expressed the hope that global models, after a long stretch of playing the odd line of evidence

275  out in assessments of global energy budget problems (Bellouin et al., 2020; Sherwood et al., 2020), might be returning to a more equal role in the balance and struggle between conflicting lines of evidence. One way in which the global model perspective

shores up the strength of the multiline assessment by providing information not available from the other lines of evidence: being able to test causality and showing that PD internal variability may not even correctly predict the sign of the causal cloud water adjustment to the anthropogenic cloud droplet perturbation.

280  Causality (or, in this case, lack of causality) is easy to establish in a model experiment but very difficult in observations. Where the noncausal correlation originates is another question that models can, in principle, answer definitively by shutting off confounding model mechanisms in mechanism-denial experiments. In part 2 of this series, we will more fully use the power of models as hypothesis testers by performing perturbed-physics and mechanism-denial experiments. In this paper, we have restricted ourselves to slice-and-dice analyses that could, in principle, also be performed on observations. We hope that

285  they will be performed on observations, especially if Lagrangian investigation of cloud life cycle (e.g., Eastman et al., 2022; Christensen et al., 2023) and observational fingerprints of loss processes (e.g., Varble et al., 2023) can be included.

Whether lack of causality in the model system implies lack of causality in the real atmosphere is a question that models alone cannot address, so we do not yet know how worried we need to be about the sign difference between correlation and causation in the model $N_d$–$\mathcal{L}$ relationship. When it comes to the non-GCM lines of evidence, one can quibble with the representativeness of the causal evidence and with causality in the representative evidence – at the very least, these model results are a flashing

290  red warning sign hanging over our interpretation of the $\mathcal{L}$ adjustment component of ERFaci.

# 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.

Albrecht, B. A.: Aerosols, Cloud Microphysics, and Fractional Cloudiness, Science, 245, 1227–1230, 1989.

Arola, A., Lipponen, A., Kolmonen, P., Virtanen, T. H., Bellouin, N., Grosvenor, D. P., Gryspeerdt, E., Quaas, J., and Kokkola, H.: Aerosol effects on clouds are concealed by natural cloud heterogeneity and satellite retrieval errors, Nature Commun., 13, 7357, https://doi.org/10.1038/s41467-022-34948-5, 2022.

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.

Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration from satellite, J. Geophys. Res., 112, D02 201, https://doi.org/10.1029/2006JD007547, 2007.

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., and Schanen, D. P.: Higher-Order Turbulence Closure and Its Impact on Climate Simulations in the Community Atmosphere Model, J. Climate, 26, 9655–9676, https://doi.org/10.1175/JCLI-D-13-00075.1, 2013.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S., Sherwood, S., Stevens, B., and Zhang, X.: Clouds and Aerosols, book section Chapter 7, pp. 571–658, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.016, 2014.

Brenguier, J. L., Pawlowska, H., Schüller, L., Preusker, R., Fischer, J., and Fouquart, Y.: Radiative properties of boundary layer clouds: Droplet effective radius versus number concentration, J. Atmos. Sci., 57, 803–821, https://doi.org/10.1175/1520-0469(2000)057<0803:RPOBLC>2.0.CO;2, 2000.

Brenguier, J. L., Pawlowska, H., and Schüller, L.: Cloud microphysical and radiative properties for parameterization and satellite monitoring of the indirect effect of aerosol on climate -: art. no. 8632, J. Geophys. Res., 108, 8632, https://doi.org/10.1029/2002JD002682, 2003.

Bretherton, C. S. and Park, S.: A New Moist Turbulence Parameterization in the Community Atmosphere Model, J. Climate, 22, 3422–3448, https://doi.org/10.1175/2008JCLI2556.1, 2009.

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.

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., Gettelman, A., Cermak, J., Dagan, G., Diamond, M., Douglas, A., Feingold, G., Glassmeier, F., Goren, T., Grosvenor, D. P., Gryspeerdt, E., Kahn, R., Li, Z., Ma, P.-L., Malavelle, F., McCoy, I. L., McCoy, D. T., McFarquhar, G., Mülmenstädt, J., Pal, S., Possner, A., Povey, A., Quaas, J., Rosenfeld, D., Schmidt, A., Schroedner, R., Sorooshian, A., Stier, P., Toll, V., Watson-Parris, D., Wood,

R., Yang, M., and Yuan, T.: Opportunistic experiments to constrain aerosol effective radiative forcing, Atmos. Chem. Phys., 22, 641–674, https://doi.org/10.5194/acp-22-641-2022, 2022.

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

350     a Lagrangian framework, Atmos. Chem. Phys., 23, 2789–2812, https://doi.org/10.5194/acp-23-2789-2023, 2023.

Comstock, K. K., Wood, R., Yuter, S. E., and Bretherton, C. S.: Reflectivity and rain rate in and below drizzling stratocumulus, Quart. J. Roy. Meteorol. Soc., 130, 2891–2918, https://doi.org/10.1256/qj.03.187, 2004.

Diamond, M. S., Director, H. M., Eastman, R., Possner, A., and Wood, R.: Substantial Cloud Brightening From Shipping in Subtropical Low Clouds, AGU Adv., 1, e2019AV000 111, https://doi.org/https://doi.org/10.1029/2019AV000111, e2019AV000111 2019AV000111, 2020.

355     Dipu, S., Schwarz, M., Ekman, A. M. L., Gryspeerdt, E., Goren, T., Sourdeval, O., Mülmenstädt, J., and Quaas, J.: Exploring satellite-derived relationships between cloud droplet number concentration and liquid water path using large-domain large-eddy simulation, Tellus, 74, 176–188, https://doi.org/10.16993/tellusb.27, 2022.

Eastman, R., McCoy, I. L., and Wood, R.: Wind, Rain, and the Closed to Open Cell Transition in Subtropical Marine Stratocumulus, J. Geophys. Res., 127, e2022JD036 795, https://doi.org/10.1029/2022JD036795, 2022.

360     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.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D., Mauritsen, T., Palmer, M., Watanabe, M., Wild,

365     M., , and Zhang, H.: The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity, chap. Chapter 7, pp. 923–1054, Cambridge University Press, https://doi.org/10.1017/9781009157896.009, 2021.

Gettelman, A.: Putting the clouds back in aerosol-cloud interactions, Atmos. Chem. Phys., 15, 12 397–12 411, https://doi.org/10.5194/acp-15-12397-2015, 2015.

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., Lamarque, J.-F., Fasullo, J. T., Bailey, D. A.,

370     Lawrence, D. M., and Mills, M. J.: High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2), Geophys. Res. Lett., 46, 8329–8337, https://doi.org/10.1029/2019GL083978, 2019.

Ghan, S., Wang, M., Zhang, S., Ferrachat, S., Gettelman, A., Griesfeller, J., Kipling, Z., Lohmann, U., Morrison, H., Neubauer, D., Partridge, D. G., Stier, P., Takemura, T., Wang, H., and Zhang, K.: Challenges in constraining anthropogenic aerosol effects on cloud radiative forcing using present-day spatiotemporal variability, Proc. Nat. Acad. Sci. USA, 113, 5804–5811, https://doi.org/10.1073/pnas.1514036113, 2016.

375     Glassmeier, F., Hoffmann, F., Johnson, J. S., Yamaguchi, T., Carslaw, K. S., and Feingold, G.: Aerosol-cloud-climate cooling overestimated by ship-track data, Science, 371, 485–+, https://doi.org/10.1126/science.abd3980, 2021.

Golaz, J.-C., Larson, V. E., Hansen, J. A., Schanen, D. P., and Griffin, B. M.: Elucidating model inadequacies in a cloud parameterization by use of an ensemble-based calibration framework, Mon. Weather Rev., 135, 4077–4096, https://doi.org/10.1175/2007MWR2008.1, 2007.

Grandey, B. S. and Stier, P.: A critical look at spatial scale choices in satellite-based aerosol indirect effect studies, Atmos. Chem. Phys., 10,

380     11 459–11 470, https://doi.org/10.5194/acp-10-11459-2010, 2010.

Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R., Cairns, B., Chiu, J. C., Christensen, M., Deneke, H., Diamond, M., Feingold, G., Fridlind, A., Huenerbein, A., Knist, C., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D., Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., van Diedenhoven, B., Wendisch, M., Werner,

F., Wood, R., Zhang, Z., and Quaas, J.: Remote Sensing of Droplet Number Concentration in Warm Clouds: A Review of the Current

385   State of Knowledge and Perspectives, Rev. Geophys., 56, 409–453, https://doi.org/10.1029/2017RG000593, 2018.

Gryspeerdt, E. and Stier, P.: Regime-based analysis of aerosol-cloud interactions, Geophys. Res. Lett., 39, L21 802, https://doi.org/
10.1029/2012GL053221, 2012.

Gryspeerdt, E., Goren, T., Sourdeval, O., Quaas, J., Mülmenstädt, J., Dipu, S., Unglaub, C., Gettelman, A., and Christensen, M.: Constraining
the aerosol influence on cloud liquid water path, Atmos. Chem. Phys., 19, 5331–5347, https://doi.org/10.5194/acp-19-5331-2019, 2019.

390   Gryspeerdt, E., Müelmenstädt, 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.

395   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.

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.

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

Larson, V. E. and Golaz, J. C.: Using probability density functions to derive consistent closure relationships among higher-order moments,
Mon. Weather Rev., 133, 1023–1042, https://doi.org/10.1175/MWR2902.1, 2005.

Malavelle, F. F., Haywood, J. M., Jones, A., Gettelman, A., Larisse, L. C., Bauduin, S., Allan, R. P., Karset, I. H. H., Kristjansson, J. E.,

405   Oreopoulos, L., Ho, N. C., Lee, D., Bellouin, N., Boucher, O., Grosvenor, D. P., Carslaw, K. S., Dhomse, S., Mann, G. W., Schmidt, A.,
Coe, H., Hartley, M. E., Dalvi, M., Hill, A. A., Johnson, B. T., Johnson, C. E., Knight, J. R., O'Connor, F. M., Partridge, D. G., Stier, P.,
Myhre, G., Platnick, S., Stephens, G. L., Takahashi, H., and Thordarson, T.: Strong constraints on aerosol-cloud interactions from volcanic
eruptions, Nature, 546, 485–491, https://doi.org/10.1038/nature22974, 2017.

Medeiros, B. and Stevens, B.: Revealing differences in GCM representations of low clouds, Clim. Dynam., 36, 385–399, https://doi.org/

410   10.1007/s00382-009-0694-5, 2011.

Morrison, H. and Gettelman, A.: A new two-moment bulk stratiform cloud microphysics scheme in the community atmosphere model,
version 3 (CAM3). Part I: Description and numerical tests, J. Climate, 21, 3642–3659, https://doi.org/10.1175/2008JCLI2105.1, 2008.

Muhlbauer, A., McCoy, I. L., and Wood, R.: Climatology of stratocumulus cloud morphologies: microphysical properties and radiative
effects, Atmos. Chem. Phys., 14, 6695–6716, https://doi.org/10.5194/acp-14-6695-2014, 2014.

415   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.,

420   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.: Can general-circulation models represent cloud adjustments to aerosol–cloud interactions?, Atmos. Chem. Phys.,
425    https://www.overleaf.com/read/mzprhdypdfmw#093585, in prep.

Painemal, D. and Zuidema, P.: Assessment of MODIS cloud effective radius and optical thickness retrievals over the Southeast Pacific with VOCALS-REx in situ measurements, J. Geophys. Res., 116, D24 206, https://doi.org/10.1029/2011JD016155, 2011.

Pawlowska, H. and Brenguier, J. L.: An observational study of drizzle formation in stratocumulus clouds for general circulation model (GCM) parameterizations, J. Geophys. Res., 108, 8630, https://doi.org/10.1029/2002JD002679, 2003.

430    Pearl, J. and Mackenzie, D.: The Book of Why: The New Science of Cause and Effect, Basic Books, New York, USA, 2018.

Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., and Hofmann, R. J. P.: Reconciling Simulated and Observed Views of Clouds: MODIS, ISCCP, and the Limits of Instrument Simulators, J. Climate, 25, 4699–4720, https://doi.org/10.1175/JCLI-D-11-00267.1, 2012.

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G. T., Zhang, Z., Hubanks, P. A., Holz, R. E., Yang, P., Ridgway, W. L., and Riedi, J.: The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples
435    From Terra and Aqua, IEEE Transact. Geosci. Remote Sens., 55, 502–525, https://doi.org/10.1109/TGRS.2016.2610522, 2017.

Possner, A., Eastman, R., Bender, F., and Glassmeier, F.: Deconvolution of boundary layer depth and aerosol constraints on cloud water path in subtropical stratocumulus decks, Atmos. Chem. Phys., 20, 3609–3621, https://doi.org/10.5194/acp-20-3609-2020, 2020.

Quaas, J., Arola, A., Cairns, B., Christensen, M., Deneke, H., Ekman, A. M. L., Feingold, G., Fridlind, A., Gryspeerdt, E., Hasekamp, O., Li, Z., Lipponen, A., Ma, P.-L., Mülmenstädt, J., Nenes, A., Penner, J. E., Rosenfeld, D., Schroedner, R., Sinclair, K., Sourdeval, O., Stier, P.,
440    Tesche, M., van Diedenhoven, B., and Wendisch, M.: Constraining the Twomey effect from satellite observations: issues and perspectives, Atmos. Chem. Phys., 20, 15 079–15 099, https://doi.org/10.5194/acp-20-15079-2020, 2020.

Quaas, J., Jia, H., Smith, C., Albright, A. L., Aas, W., Bellouin, N., Boucher, O., Doutriaux-Boucher, M., Forster, P. M., Grosvenor, D., Jenkins, S., Klimont, Z., Loeb, N. G., Ma, X., Naik, V., Paulot, F., Stier, P., Wild, M., Myhre, G., and Schulz, M.: Robust evidence for reversal in the aerosol effective climate forcing trend, Atmos. Chem. Phys., 22, 12 221–12 239, https://doi.org/10.5194/acp-22-12221-
445    2022, 2022.

Rasch, P. J., Xie, S., Ma, P.-L., Lin, W., Wang, H., Tang, Q., Burrows, S. M., Caldwell, P., Zhang, K., Easter, R. C., Cameron-Smith, P., Singh, B., Wan, H., Golaz, J.-C., Harrop, B. E., Roesler, E., Bacmeister, J., Larson, V. E., Evans, K. J., Qian, Y., Taylor, M., Leung, L. R., Zhang, Y., Brent, L., Branstetter, M., Hannay, C., Mahajan, S., Mametjanov, A., Neale, R., Richter, J. H., Yoon, J.-H., Zender, C. S., Bader, D., Flanner, M., Foucar, J. G., Jacob, R., Keen, N., Klein, S. A., Liu, X., Salinger, A. G., Shrivastava, M., and Yang, Y.:
450    An Overview of the Atmospheric Component of the Energy Exascale Earth system Model, J. Adv. Model. Earth Syst., 11, 2377–2411, https://doi.org/10.1029/2019MS001629, 2019.

Rossow, W. B., Tselioudis, G., Polak, A., and Jakob, C.: Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures, Geophys. Res. Lett., 32, L21 812, https://doi.org/10.1029/2005GL024584, 2005.

Salzmann, M., Ming, Y., Golaz, J.-C., Ginoux, P. A., Morrison, H., Gettelman, A., Kraemer, M., and Donner, L. J.: Two-moment bulk
455    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.

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., von der Heydt, A. S., Knutti, R., Mauritsen,

T., Norris, J. R., Proistosescu, C., Rugenstein, M., Schmidt, G. A., Tokarska, K. B., and Zelinka, M. D.: An Assessment of Earth's Climate
460    Sensitivity Using Multiple Lines of Evidence, Rev. Geophys., 58, e2019RG000 678, https://doi.org/10.1029/2019RG000678, 2020.

Simpson, E.: The Interpretation of Interaction in Contingency Tables, Journal of the Royal Statistical Society Series B-Statistical Methodol-
       ogy, 13, 238–241, 1951.

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., and Haynes, J.: Dreary
       state of precipitation in global models, J. Geophys. Res., 115, D24 211, https://doi.org/10.1029/2010JD014532, 2010.

465   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.

Swales, D. J., Pincus, R., and Bodas-Salcedo, A.: The Cloud Feedback Model Intercomparison Project Observational Simulator Package:
       Version 2, Geosci. Model Dev., 11, 77–81, https://doi.org/10.5194/gmd-11-77-2018, 2018.

Toll, V., Christensen, M., Quaas, J., and Bellouin, N.: Weak average liquid-cloud-water response to anthropogenic aerosols, Nature, 572,
470    51–+, https://doi.org/10.1038/s41586-019-1423-9, 2019.

Twomey, S.: Influence of pollution on shortwave albedo of clouds, J. Atmos. Sci., 34, 1149–1152, https://doi.org/10.1175/1520-
       0469(1977)034<1149:TIOPOT>2.0.CO;2, 1977.

Unglaub, C., Block, K., Mülmenstädt, J., Sourdeval, O., and Quaas, J.: A new classification of satellite-derived liquid water cloud regimes at
       cloud scale, Atmos. Chem. Phys., 20, 2407–2418, https://doi.org/10.5194/acp-20-2407-2020, 2020.

475   Varble, A. C., Ma, P.-L., Christensen, M. W., Mülmenstädt, J., Tang, S., , and Fast, J.: Evaluation of Liquid Cloud Albedo Suscepti-
       bility in E3SM Using Coupled Eastern North Atlantic Surface and Satellite Retrievals, Atmos. Chem. Phys. Discuss., https://doi.org/
       10.5194/egusphere-2023-998, 2023.

Zhang, S., Wang, M., Ghan, S. J., Ding, A., Wang, H., Zhang, K., Neubauer, D., Lohmann, U., Ferrachat, S., Takeamura, T., Gettelman, A.,
       Morrison, H., Lee, Y., Shindell, D. T., Partridge, D. G., Stier, P., Kipling, Z., and Fu, C.: On the characteristics of aerosol indirect effect
480    based on dynamic regimes in global climate models, Atmos. Chem. Phys., 16, 2765–2783, https://doi.org/10.5194/acp-16-2765-2016,
       2016.

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., Chen, J.-H., Chen, X., Donner, L. J., Dunne, J. P., Dunne, K., Durachta,
       J., Fan, S.-M., Freidenreich, S. M., Garner, S. T., Ginoux, P., Harris, L. M., Horowitz, L. W., Krasting, J. P., Langenhorst, A. R., Liang, Z.,
       Lin, P., Lin, S.-J., Malyshev, S. L., Mason, E., Milly, P. C. D., Ming, Y., Naik, V., Paulot, F., Paynter, D., Phillipps, P., Radhakrishnan, A.,
485    Ramaswamy, V., Robinson, T., Schwarzkopf, D., Seman, C. J., Shevliakova, E., Shen, Z., Shin, H., Silvers, L. G., Wilson, J. R., Winton,
       M., Wittenberg, A. T., Wyman, B., and Xiang, B.: The GFDL Global Atmosphere and Land Model AM4.0/LM4.0:2. Model Description,
       Sensitivity Studies, and Tuning Strategies, J. Adv. Model. Earth Syst., 10, 735–769, https://doi.org/10.1002/2017MS001209, 2018.
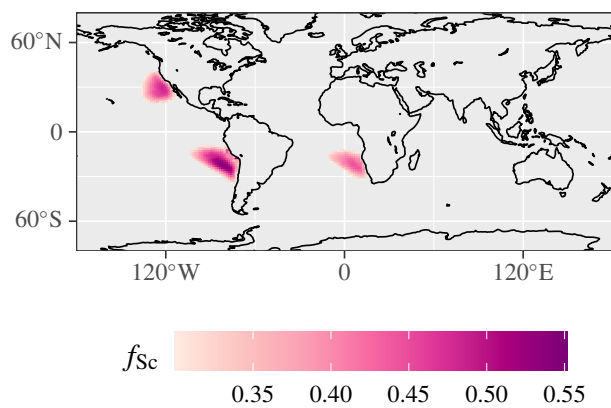
**Figure 1.** Occurrence fraction $f_{Sc}$ of Sc conditions by the Medeiros and Stevens (2011) criteria in EAMv1, shown where $f_{Sc} > 0.3$ (NEP, southeast Pacific, southeast Atlantic Sc regions).
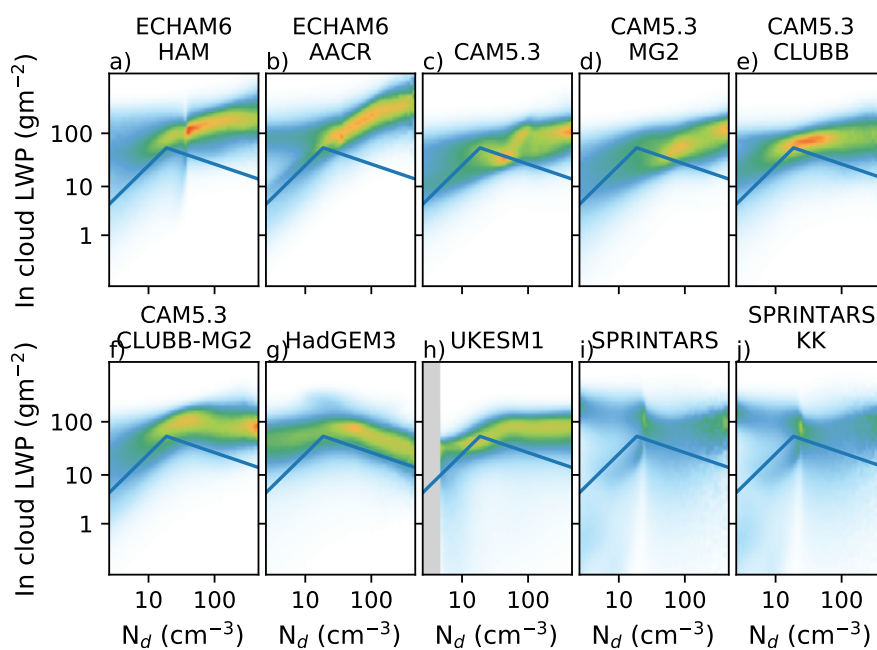
**Figure 2.** AeroCom IND3 state-of-the-art models' $N_d$–$\mathcal{L}$ relationship. The satellite inverted v relationship (Gryspeerdt et al., 2019) is indicated by the solid line.
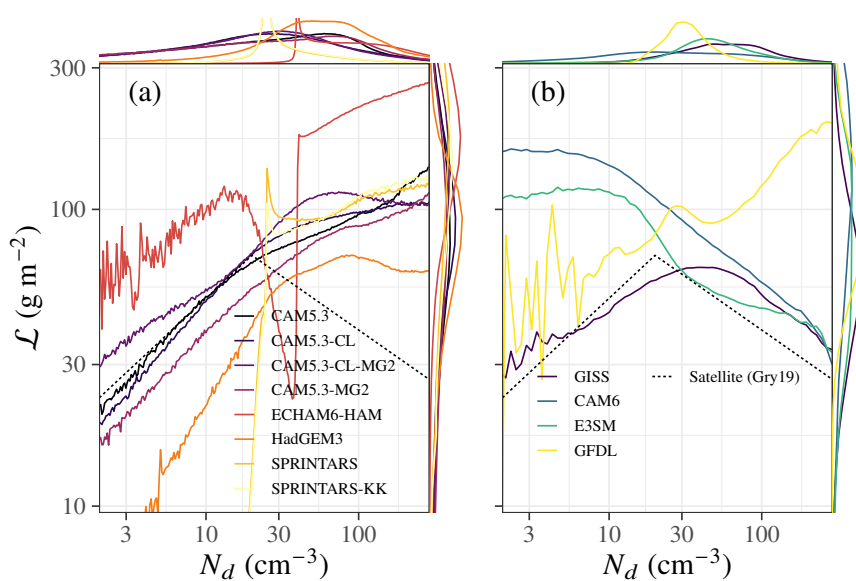
**Figure 3.** AeroCom IND3 state-of-the-art models' marginal distributions of $N_d$ and $\mathcal{L}$ and $N_d$–$\mathcal{L}$ relationship (a) compared to the CMIP6-era state-of-the-art models' $N_d$–$\mathcal{L}$ relationship (b). The satellite inverted v relationship (Gryspeerdt et al., 2019) is indicated by the dotted line. Three of the four CMIP6 models examined are qualitatively similar to the satellite result in the sense that the $N_d$–$\mathcal{L}$ correlation turns negative at moderate $N_d$.
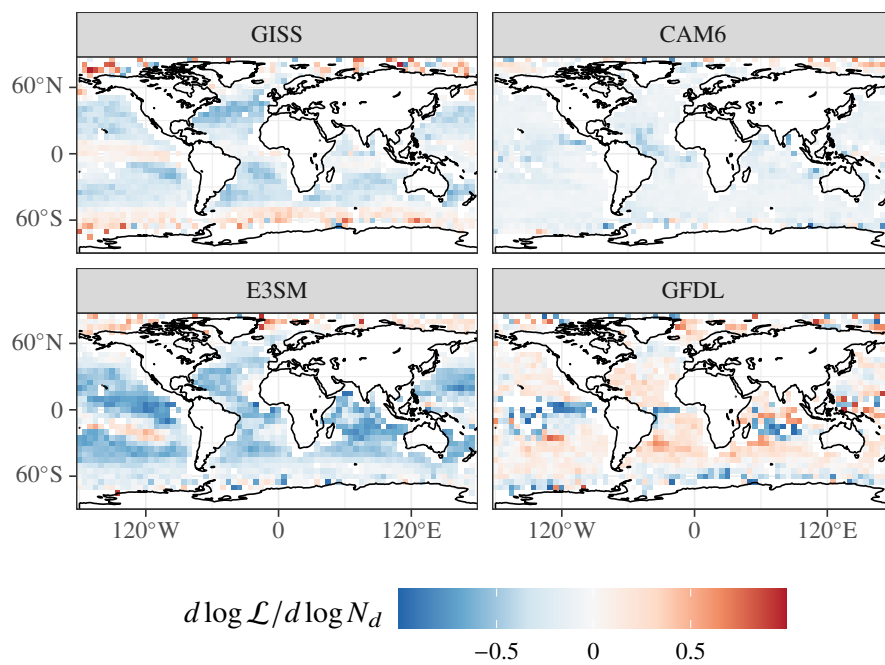
**Figure 4.** Geographic distribution of $d \log \mathcal{L} / d \log N_d$. Model output is aggregated to $5° \times 5°$ latitude–longitude boxes before calculating linear regression slopes of $\log \mathcal{L}$ against $\log N_d$.
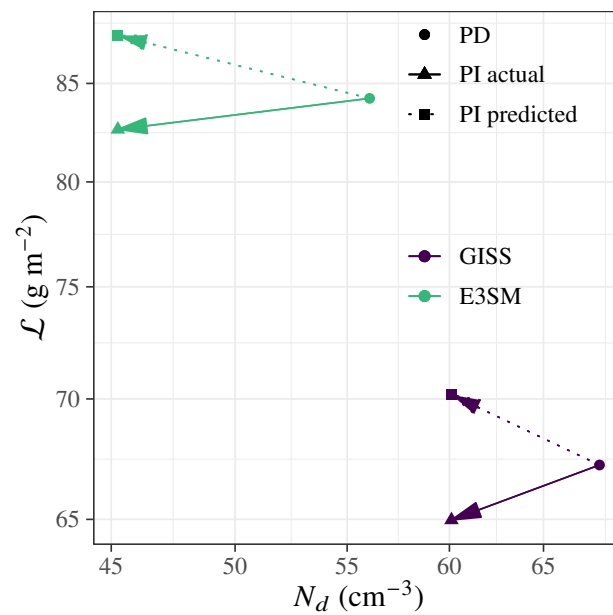
**Figure 5.** PI–PD $\mathcal{L}$ change from the causal experiment (solid arrow) contrasted with the change predicted from the PD internal variability (dashed arrow). The mean $\log\mathcal{L}$ as a function of $\log N_d$ from the PD model run (Fig. 3) is used to predict PI mean $\log\mathcal{L}$ from the PI $\log N_d$ distribution. Even though the PD $N_d$–$\mathcal{L}$ correlation is negative, $\mathcal{L}_{\text{PD}} > \mathcal{L}_{\text{PI}}$.
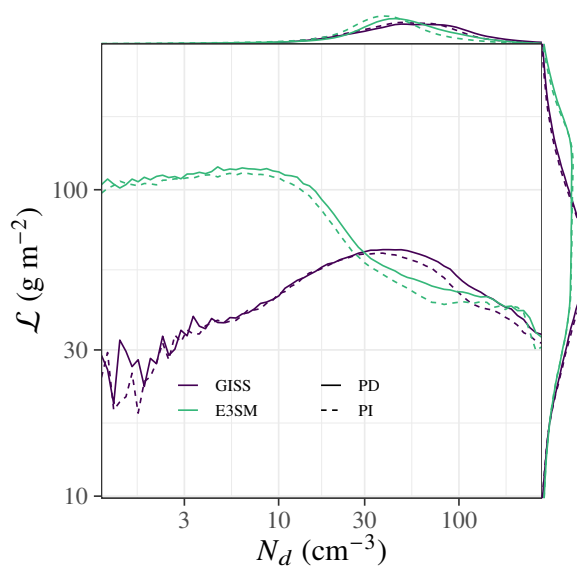
**Figure 6.** PD (solid) and PI (dashed) $N_d$ and $\mathcal{L}$ marginal distributions and $N_d$–$\mathcal{L}$ correlation in two GCMs with unrelated turbulence schemes. The $N_d$–$\mathcal{L}$ relationship based on internal variability within one climate state is not universal across the states.
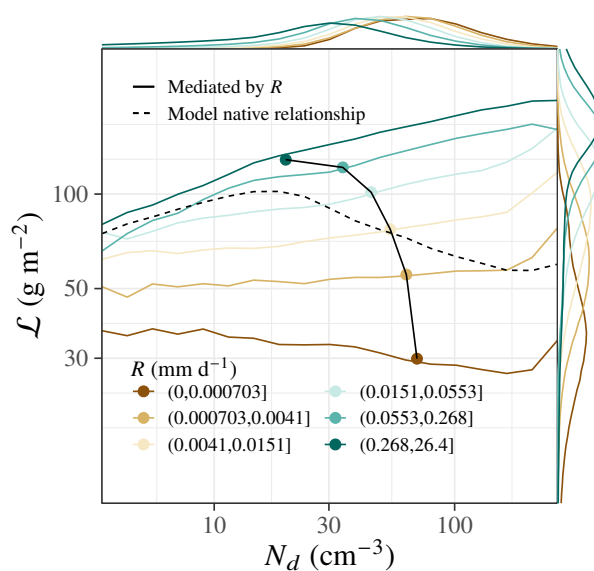
**Figure 7.** Precipitation-stratified $N_d$ and $\mathcal{L}$ marginal distributions and $N_d$–$\mathcal{L}$ relationships (colored lines). The dashed black line shows the unstratified $N_d$–$\mathcal{L}$ relationship. The solid black line connects the mean $(N_d, \mathcal{L})$ in each precipitation sextile (colored dots). Binning by precipitation intensity exposes a precipitation-mediated negative $N_d$–$\mathcal{L}$ covariability with a much steeper slope than the overall $N_d$–$\mathcal{L}$ correlation, even though the $N_d$–$\mathcal{L}$ correlation within all but the least precipitating sextile is positive.
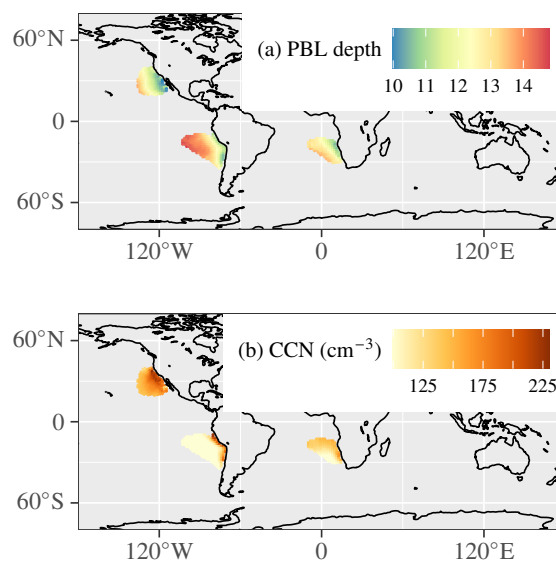
**Figure 8.** Within the Sc regime, synoptic-scale meteorology results in strong spatial covariability of (a) PBL depth (measured in model levels) and (b) CCN concentration at 0.2% supersaturation averaged over the depth of the PBL. Both of these variables are functions of airmass continentality.
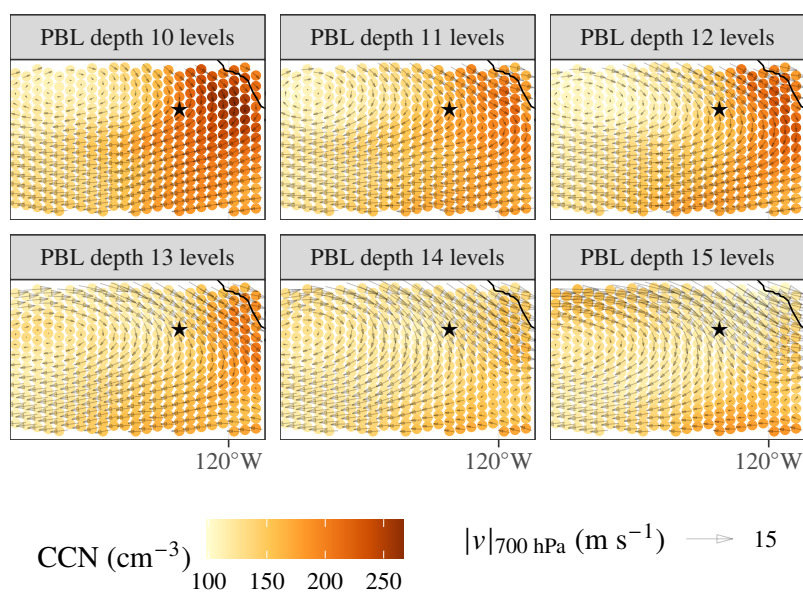
**Figure 9.** Within the Sc regime, synoptic-scale meteorology results in strong temporal covariability between CCN and PBL depth. This is exemplified by the NEP; the Southern California Bight is depicted in top right corner. The star indicates the grid point with the highest occurrence fraction of Sc conditions according to the criteria of Medeiros and Stevens (2011). (Aggregates exclude points over land.)
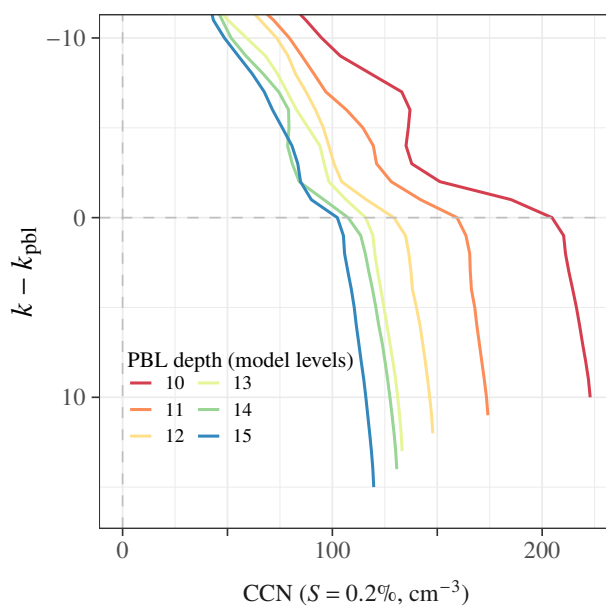
**Figure 10.** Temporal and regional-mean CCN concentration profiles in the NEP Sc region stratified by PBL depth. Within the Sc regime, CCN is strongly sorted by PBL depth, illustrating the strong covariability between PBL thermodynamic structure and aerosol advection. The central 90% of the PBL depth range (between 10 and 15 model levels, corresponding approximately to between 750–1400 m) is shown to avoid outliers in the low-statistics PBL depth bins. The PBL depth is measured in units of model levels $k$, with $k$ decreasing downward from the level of the PBL-capping inversion $k_{pbl}$ to the model level closest to the surface ($k = 72$ in EAMv1).
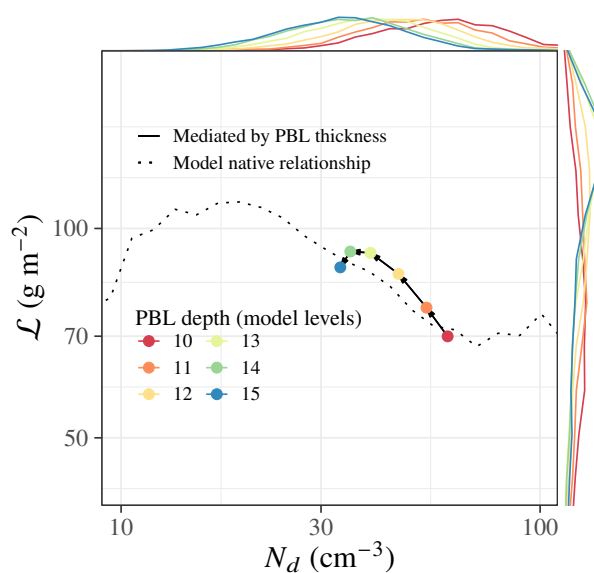
**Figure 11.** Within the Sc regime, PBL depth–CCN covariability leads to a negative $N_d$–$\mathcal{L}$ correlation with slope similar to the overall $N_d$–$\mathcal{L}$ correlation. The dashed black line shows the $N_d$–$\mathcal{L}$ relationship not stratified by PBL depth. The solid black line connects the mean $(N_d, \mathcal{L})$ at each PBL depth (colored dots). The central 90% of the PBL depth range (between 10 and 15 model levels) is shown to avoid outliers in the low-statistics PBL depth bins.
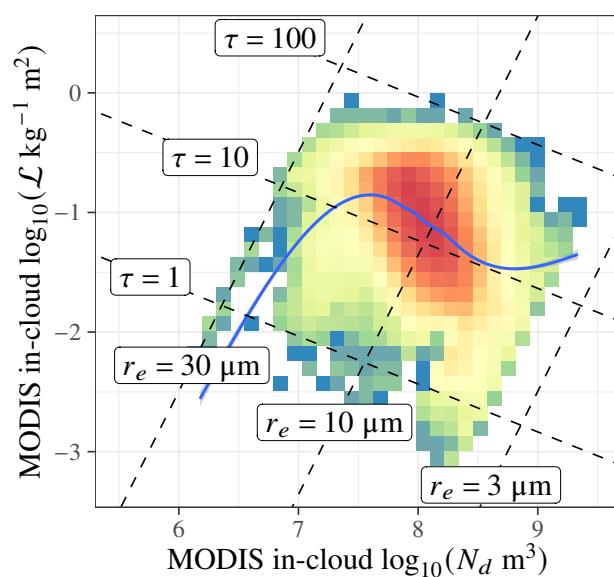
**Figure 12.** Joint probability distribution of E3SM MODIS-simulated $\mathcal{L}$ and $N_d$. Isolines of $r_e$ and $\tau$, from which adiabatic $N_d$ and $\mathcal{L}$ are retrieved, are overlaid. The mean $\log \mathcal{L}$ as a function of $\log N_d$ is shown as a blue line. Because the model imposes a rectangular limit on $r_e$ and $\tau$, the $N_d$–$\mathcal{L}$ phase space has parallelogram-shaped boundaries. At least part of the rise, fall, and repeated rise of $\mathcal{L}$ as a function of $N_d$ (blue line) is due to these phase-space boundaries cutting off the upper and lower parts of the $\mathcal{L}$ distribution.

**Table 1.** GCM experiments used in this analysis.

| Experiment | Emissions | Cloud selection | Duration |
|---|---|---|---|
| Multimodel PD | Present-day | Global overcast | 2010, nudged |
| Multimodel PI | Preindustrial | Global overcast | 2011, nudged |
| E3SM + COSP simulator | Present-day | NEP Sc regime | 2010, nudged |
| E3SM precip | Present-day | Sc regime | 2010–2014, nudged |