

Aircraft In-situ Measurements from SOCRATES Constrain the Anthropogenic Perturbations of Cloud Droplet Number

Ci Song¹, Daniel T. McCoy¹, Isabel L. McCoy^{2, 3}, Hunter Brown¹, Andrew Gettelman⁴, Trude Eidhammer⁵, and Donifan Barahona⁶

¹Department of Atmospheric Science, University of Wyoming, Laramie, WY, USA

²Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

³National Oceanic and Atmospheric Administration, Chemical Sciences Laboratory, Boulder, CO, USA

⁴Pacific Northwest National Laboratory, Richland, WA, USA

⁵NSF National Center for Atmospheric Research, Boulder, CO, USA

⁶Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Correspondence: Ci Song (csong@uwyo.edu)

Abstract. Aerosol-cloud interactions (ACI) in warm clouds alter reflected shortwave radiation by influencing cloud microphysical and macrophysical properties. The variable of state controlling ACI is the cloud droplet number concentration (N_d). Here, we examine the perturbations in N_d due to anthropogenic aerosols ($\Delta N_{d,PD-PI}$) using a perturbed parameter ensemble (PPE) hosted in the sixth Community Atmosphere Model (CAM6). Surrogate models are created for the CAM6 PPE outputs and are used to generate 1 million model variants of CAM6 by sampling 45 sources of parameter uncertainty. The range of uncertain physical parameters related to ACI are constrained with observations of aerosol and cloud properties from SOCRATES. The likely range of uncertain parameters and the associated range of $\Delta N_{d,PD-PI}$ are more strongly constrained with observations of N_d relative to observations of cloud condensation nuclei. We conduct sensitivity tests of how constraints on $\Delta N_{d,PD-PI}$ are affected by systematic uncertainties in observations and our limitations in our surrogate models created for CAM6 PPE outputs. Based on this, we provide guidance on the impact of reducing systematic uncertainty in airborne microphysical observations and in surrogate models.

1 Introduction

Clouds play an essential role in setting Earth's top of atmosphere energy flux by reflecting incoming shortwave radiation back to space. Aerosols are important for cloud formation as they serve as cloud condensation nuclei (CCN) for water vapor to condense onto. CCN make cloud droplet formation possible in atmospheric conditions. Aerosols from anthropogenic emissions alter cloud droplet number concentration (N_d) by acting as CCN, enhancing cloud reflectivity (Twomey, 1977). The change in reflected shortwave radiation (i.e., radiative forcing: RF in W/m²) through changes in N_d is referred to as the instantaneous radiative forcing due to aerosol-cloud interactions (IRFaci). According to the formulation in Bellouin et al. (2020), IRFaci is given by

$$IRFaci = \left. \frac{\partial R}{\partial \ln N_d} \right|_{LWPe, C} \cdot \Delta \ln N_d \quad (1)$$

20 where R is the net radiative flux. LWP_c is the in-cloud liquid water path (LWP) and $\Delta \ln N_d$ is the fractional perturbation in N_d (Ghan et al., 2016; Bellouin et al., 2020). The vertical line in the partial derivative denotes LWP_c and cloud fraction (C) are held constant (Bellouin et al., 2020). With changes in N_d , cloud macrophysical properties can be altered in response to changes in cloud microphysics, such as cloud lifetime, liquid water content and cloud cover (Ackerman et al., 2004). The RF caused by modifications to cloud macrophysics is referred to as aerosol-cloud adjustment and is given by

$$RF_{adjustment} = \left(\frac{\partial R}{\partial C} \frac{dC}{d \ln N_d} + \frac{\partial R}{\partial LWP_c} \frac{dLWP_c}{d \ln N_d} \right) \cdot \Delta \ln N_d \quad (2)$$

25 The sum of radiative forcing from IRFaci and aerosol-cloud adjustment is termed effective RF due to ACI (ERFaci), which can be expressed as

$$ERFaci = \left(\frac{\partial R}{\partial \ln N_d} \Big|_{LWP_c, C} + \frac{\partial R}{\partial C} \cdot \frac{dC}{d \ln N_d} + \frac{\partial R}{\partial LWP_c} \cdot \frac{dLWP_c}{d \ln N_d} \right) \cdot \Delta \ln N_d \quad (3)$$

Recent assessments place ERFaci as the largest uncertainty in anthropogenic climate forcing. This uncertainty also complicates efforts to infer climate sensitivity from the historical record, as the cooling from ACI can mask the warming effects of greenhouse gases (GHGs) (Bellouin et al., 2020; Forster, 2016; Watson-Parris and Smith, 2022).

30 Earth system models (ESMs) are essential for estimating ERFaci as they can estimate the unobservable preindustrial baseline of the atmosphere (Carslaw et al., 2017; Wall et al., 2022). However, ESMs are uncertain in their representations of aerosols and their climate effects. This uncertainty can be related to structural uncertainty (what processes to include in a model) (Regayre et al., 2023) and parametric uncertainty (how the values of parameters in the mathematical representation of processes are set in the ESM) (Regayre et al., 2018). The uncertainties in ERFaci related to parametrizations of unresolved aerosol processes, emissions, and cloud microphysical processes within a single model can be as large as the spread across models with different model structures. This supports the utility of understanding parametric uncertainties (Johnson et al., 2018). A commonly used method is to employ a perturbed parameter ensemble (PPE). This method involves exploring many possible parameter combinations across their uncertainty range to quantify the range of possible outcomes. The plausible range of ERFaci can be estimated using a set of parameter combinations, provided there is good agreement between observations and the model simulations generated by those parameter combinations (Regayre et al., 2018).

Wood (2012) argues that the variable of state (or most important variable) in understanding ACI is the N_d . Effectively, changes in N_d play a pivotal role in governing cloud radiative and macrophysical behavior. This means that to reduce uncertainty in ERFaci, constraining the anthropogenic perturbation to N_d is essential as both IRFaci (Eq 1) and aerosol-cloud adjustment (Eq 2) scale with change in N_d (Bellouin et al., 2020; Song et al., 2024).

45 One obstacle in seeking an observational constraint on the N_d response to anthropogenic aerosol is that the processes driving the N_d response primarily occur at the microscale and the result of these processes poses observational challenges. Past studies have used observations of N_d from spaceborne remote sensing to constrain the change in N_d during historical periods, achieving consistent observational constraints across different host models using the same observations (McCoy et al., 2020; Song et al., 2024; Gryspeerdt et al., 2016). However, observations of aerosol and cloud microphysical properties from remote sensing are known to have uncertainties arising from factors such as assumptions about particle size distributions, cloud

microphysics, and radiative transfer models used in the retrieval process (Grosvenor et al., 2018; Zhang et al., 2016; Gryspeerdt et al., 2022).

In-situ measurements provide direct measurements of aerosol and cloud microphysical properties without reliance on retrieval algorithms or assumptions used in remote sensing. It also measures more detailed microphysical properties such as aerosol size distribution, chemical composition, and cloud droplet number concentration and size distributions. However, in-situ measurements can suffer from a wide variety of instrument biases and limitations and the impact of these limitations on our ability to use them for climate studies is not well characterized. For instance, instruments used for measuring aerosol and cloud properties can only detect subsets of the full particle distribution due to their limited sampling volume, and they cannot measure the full spectrum of particle sizes (Lance et al., 2010). In-situ measurements from aircraft occur with a much smaller footprint than a typical ESM and are often targeted towards features that make them not representative to compare to an ESM grid cell (Field and Furtado, 2016). Additionally, by their nature aircraft campaigns have minimal global coverage and it is unclear how effective a constraint on global model behavior they provide.

In this paper, we focus on characterizing an observational constraint on the change in N_d during the historical period ($\Delta N_{d,PD-PI}$) based on in-situ measurements from a single campaign to illustrate the utility of combining two key tools: ESMs and airborne observations of microscale properties. We expand on previous work (Gettelman et al., 2020) by examining parametric uncertainty across a single ESM (i.e. using a PPE) and characterizing what we can learn from an airborne campaign and expanding on previous PPE work leveraging surface observations of aerosol properties (Regayre et al., 2020). We use observations of both aerosol and cloud properties from aircraft in-situ measurements. We address the following question: 1) do aerosol or cloud measurements better constrain global cloud microphysical behavior? 2) can sparse in-situ measurements produce constraints on cloud microphysical behavior on a global scale? 3) how sensitive is the observational constraint on $\Delta N_{d,PD-PI}$ to observation uncertainties? We provide this analysis with the goal of (i) showing the connection between in-situ measurements and our understanding of climate (Regayre et al., 2020) and (ii) characterizing where to expend effort in terms of sampling with in-situ measurements and model development.

2 Materials and Methods

2.1 The CAM6 Perturbed parameter ensemble

We use the Community Atmosphere Model version 6 (CAM6), which is the atmosphere component of the Community Earth System Model version CESM-2.0 (Danabasoglu et al., 2020). The CAM6 model uses a two-moment microphysics scheme for stratiform clouds, with liquid, ice, rain, and snow hydrometeors calculated as prognostic variables, allowing CAM6 to explicitly represent the aerosol indirect effect (Gettelman and Morrison, 2015; Gettelman et al., 2015).

We leverage a perturbed parameter ensemble (PPE) hosted in CAM6 (Eidhammer et al., 2024). A PPE is a large set of simulations based on the structure of a single ESM (e.g., CAM6) with a different combination of parameter values to examine parameter uncertainty (Lee et al., 2011; Carslaw et al., 2013). The CAM6 PPE is fully described in Eidhammer et al. (2024). CAM6 is run at the standard resolution of $1.25^\circ \times 0.9375^\circ$ resolution. Briefly, 262 model simulations (i.e., 262 parameter com-

binations) of CAM6 sample 45 sources of uncertainty in the parameterizations for cloud, precipitation, convection, boundary
85 layer, and aerosol processes. The 45 parameters are simultaneously perturbed using Latin Hypercube within the plausible range
of realistic values based on expert-elicitation. We examine 203 ensemble members out of 262 integrated. The remaining 59
members were excluded based on criterion: (1) the linear regression slope of N_d to CCN in log space ($d\ln N_d/d\ln CCN$) is
less than 0; (2) the correlation coefficient between N_d and CCN is less than 0.3. The two criteria are used to exclude PPE
members that are too far outside the observational constraint behaviors (i.e., the Southern Ocean field campaign measurements
90 analyzed in Figure 14 in McCoy et al. (2021)). Following Song et al. (2024), we also exclude PPE members that simulate too
much ice in tropics, which is inconsistent with satellite observations (King et al., 2013).

2.2 Model Configuration

Two scenarios are simulated and each of them use the same parameter combinations - consistent with previous studies (Song
et al., 2024). First, 2-year global simulations saved at monthly-mean are completed for pre-industrial (PI) and present-day (PD)
95 emissions. PI and PD aerosol emission scenarios are integrated from 2019 to 2020 so anthropogenic perturbations in N_d can be
calculated over global coverage by taking the difference between PI and PD. The atmosphere is nudged to horizontal winds and
temperature and sea surface temperature and sea ice fraction are prescribed from observations. Wind and temperature fields are
nudged to the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2) reanalysis (Bosilovich
et al., 2015) with 24-hour relaxation time. MERRA2 output is interpolated to CAM6 vertical resolution with standard 32
100 vertical levels from the surface to 3 hPa following Gettelman et al. (2020). Previous studies have shown the CAM6 PPE
produces a wide range of perturbations in cloud microphysics (e.g., $\Delta N_{d,PD-PI}$) and cloud macrophysics (ΔLWP_{PD-PI}).
In this study, we focus on diagnosing the parametric effects on cloud microphysical responses to anthropogenic aerosols from
different parameter combinations using the PPE.

In addition to the two-year integrations of the PPE used to calculate anthropogenic perturbations in N_d , the PPE is in-
105 tegrated over short periods consistent with the Southern Ocean Clouds, Radiation, Aerosol, Transport Experimental Study
(SOCRATES) field campaign based from Hobart, Tasmania (McFarquhar et al., 2021) (Figure 1). The SOCRATES campaign
occurred over the midlatitude Southern Ocean (SO) during austral summer and was dominated by a series of frontal systems,
postfrontal stratocumulus decks, and cyclonic activity typical of the storm track region (McFarquhar et al., 2021). Model out-
puts are saved along flight tracks over SOCRATES and is sampled at 1 min resolution following Gettelman et al. (2020). It
110 applies atmospheric nudging to horizontal winds and temperature, consistent with global simulation with PI and PD aerosol
emissions scenarios, but nudged to the period of January-March 2018 when the aircraft observations were conducted. The be-
havior of the default parameter configuration in CAM6 has been characterized using this approach in Gettelman et al. (2020);
McCoy et al. (2021); McCluskey et al. (2023); Zhou et al. (2021).

Previous studies have shown that the CAM6 PPE, configured with 2-year global simulations, produces a wide spread in
115 present-day (PD) cloud microphysical (N_d) and macrophysical (LWP) properties. The mean-state PD values have been shown
to fall within the observational range derived from satellite remote sensing (Song et al., 2024). Additionally, CAM6 simulations
along flight tracks using the default parameter configuration reproduce many features of in-situ observations, including cloud

phase, cloud location, and boundary layer structure (Gettelman et al., 2020). These results give us confidence that at least some members of the nudged PPE simulations provide a physically plausible baseline in terms of cloud microphysical and macrophysical properties. In this study, we focus specifically on microphysical properties.

2.3 Aircraft Sampling

We examine in-situ airborne observations taken from SOCRATES as our observational constraint (McFarquhar et al., 2021). The importance of the Southern Ocean (SO) to understanding the global anthropogenic contribution to N_d has been shown in several previous studies (Carslaw et al., 2013; McCoy et al., 2020). The National Science Foundation Gulfstream-V (GV) aircraft was deployed during January-March 2018 for SOCRATES. There were 15 flights sampling data from 42°S to 62°S with aerosol and cloud properties sampled at 1 Hz frequency. The GV was equipped with a variety of sensors and instruments. In this work, N_d from the cloud droplet probe (CDP) and aerosol number concentrations from the ultra-high sensitivity aerosol spectrometer (UHSAS) are examined. We focus on accumulation mode aerosols, with diameters ranging from 0.1 μm to 1 μm , reported as UHSAS100 in this paper following McCoy et al. (2021). Accumulation mode aerosol usually accounts for most of the surface area of aerosols and is a good estimate of the CCN concentration for stratocumulus updraft velocities (Seinfeld and Pandis, 2016).

With a focus on low-level, liquid cloud, we restrict the aircraft measurements of aerosol and cloud to be below 2 km. As in previous studies (McCoy et al., 2021), in-situ aircraft aerosol measurements are discarded when the liquid water content (LWC) from the CDP exceeds 0.001 g/m^{-3} , along with the subsequent 10 seconds after cloud detection. This is to avoid measurement contamination from cloud (McCoy et al., 2021). In-cloud N_d measurements are restricted to regions where the LWC from the CDP is greater than a threshold (0.1 g/m^{-3}) following McCoy et al. (2021). Because the observations of aerosols and in-cloud N_d that are considered valid for use are taken at different locations, direct comparison is challenging due to inconsistencies in spatial and temporal coverage. To make comparisons between N_d and aerosol observations, we bin the aircraft measurements by 2 min in duration and 50 m in altitude so that aerosols and N_d can be compared in the same bin. Only bins with at least ten 1 Hz flight observations are considered valid composites for use. Median values of aerosol concentration and N_d are computed for each bin for observations from each flight following McCoy et al. (2021). The instrument limitation inevitably forces us to look either at small clouds or cloud edges, where both the measurements of aerosol and cloud are valid for use. This has minimal impact on our comparison between models and observations as we colocate model output with observations as detailed in Section 2.4.

2.4 Comparison Between Model data and Observations

The default configuration of CAM6 has been extensively evaluated in Gettelman et al. (2020) and McCoy et al. (2021) and has been shown to be able to reproduce many features consistent with in-situ observations in Gettelman et al. (2020). Here, we examine a PPE that is hosted in the same model evaluated in previous studies (McCoy et al., 2021; Gettelman et al., 2020). The CAM6 model parameterization and the prior distribution of parameters (i.e., 217 sets of parameter combinations) in the PPE (Eidhammer et al., 2024) produce simulated aerosol and cloud properties that we can compare with observations to evaluate

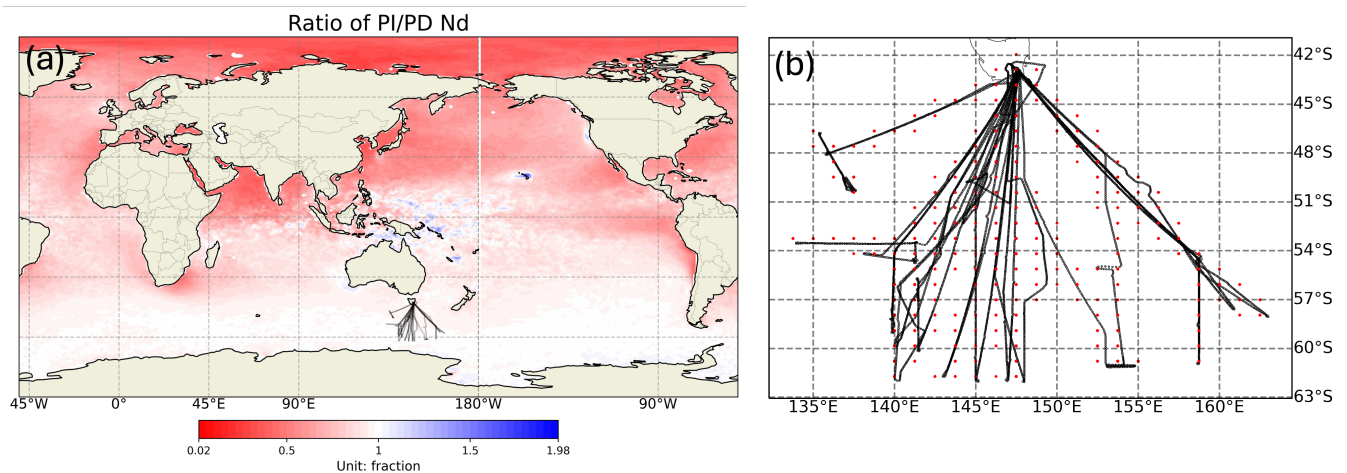


Figure 1. Maps of SOCRATES mission flight tracks from the NSF G-V aircraft. (a) Location of the SOCRATES aircraft sampling and the ratio of preindustrial to present day N_d shown in colors. The ratio is computed as $\frac{PIND}{PDNd}$ using the preindustrial and present-day simulations run for two years configured with default CAM6 parameter setting. Ratios less than 1 indicate anthropogenically polluted regions. (b) Comparison of sampling of aircraft measurements (black line) with CAM6 grid point centers (red dots). Along-flight-track simulations are run for January–March 2018, covering late austral summer into early autumn.

how process representation impacts aerosol-cloud interactions. Here, we focus on microphysical quantities that are available from in-situ measurements but hard to observe from spaceborne remote sensing.

N_d is directly available from both CAM6 and in-situ measurements from the CDP. CAM6 in-cloud N_d is calculated as N_d divided by liquid cloud fraction (when cloud fraction $\leq 1\%$, we set $N_d = 0$). This cloud fraction threshold is smaller than the one used in McCoy et al. (2021) as we found it retains more flight composites but does not significantly change the results of our analysis (Figure S1).

CCN is a subset of aerosols that can be activated to cloud droplets at a given supersaturation. CAM6 outputs CCN at a set of fixed supersaturations. Here, we look at supersaturation at 0.2%. It is found that observed CCN at 0.2% supersaturation (CCN02) has an one-to-one relationship with accumulation mode aerosol (e.g., UHSAS100) measured over SOCRATES (McCoy et al., 2021). Following previous work (McCoy et al., 2021), we use UHSAS100 as a proxy to CCN02 over SOCRATES as UHSAS100 lies very close to the one-to-one line with CCN02. This supersaturation level is shown to be representative of marine low-level stratocumulus (Hudson and Svensson, 1995).

To make comparisons between the modeled and observed N_d and aerosol properties, model data are collocated to observations by linearly interpolating to temporal and spatial locations from the $2 \text{ min} \times 50 \text{ m}$ observational composites following McCoy et al. (2021). Our comparison between observations and models follows two strategies. First, model outputs (CCN and N_d) are confronted with in-situ observations for collocated bins (flight track composites) along flight tracks for each simulation ensemble. This method allows for the evaluation of simulated CCN, N_d , and the inferred efficiency of aerosol activation ($\frac{dN_d}{dCCN}$) relative to observations within individual PPE members. The results are discussed in Section 3.1. Second,

campaign-means of CCN and N_d are calculated for each PPE ensemble and compared with campaign-means of CCN and N_d calculated from in-situ observations. In this approach, we evaluate aerosol activation efficiency across the CAM6 PPE members (run with different parameter sets) and use campaign-means of in-situ aerosol properties and N_d to constrain the CAM6 PPE (Section 3.3.3).

The intention of taking the campaign-mean is to reduce random error by averaging over a large number of samples. However, there remains potential sources of systematic error. One possible source of systematic error is from differences in sampling between the observations and the model (e.g. if the pilot only flew through clear air and avoided cloud). Sampling during airborne campaigns may have some systematic sampling biases as discussed in Field and Furtado (2016). Output from CAM6 is representative of an average within the grid box of the model, whereas flight patterns in a similar-sized domain may not be sampling randomly (e.g. focusing on convective cores). We believe that this is a minimal concern for SOCRATES. The SOCRATES flight pattern was designed to focus on cold sectors of cyclones and synoptically uplifted aerosol layers, but followed a random sampling pattern in those large-scale features (McCoy et al., 2021). In addition to any systematic errors from sampling strategy, instrument error inherent in the CDP introduces additional uncertainties in the measurements of N_d . CDP measures cloud droplets within a specific size range (i.e., 2 to 50 μm in diameter). It has limitations regarding droplets that fall outside its designed size range. Coincidence errors may occur when multiple droplets pass through the sensor’s detection volume but is counted as a single droplet. The impact of observational uncertainty on the model constraints is examined in section 3.4.

Another potential source of systematic uncertainty may arise from the use of UHSAS100 as a proxy to CCN02 over SOCRATES. While a near one-to-one relationship between UHSAS100 and CCN02 has been reported for the SOCRATES campaign (e.g., (McCoy et al., 2021)), the campaign-mean ratio of CCN02 to UHSAS100 is approximately 1.08 (± 0.3) according to their Figure S2. This suggests that UHSAS100 may underestimate CCN02 by 8% on average. Moreover, the activation diameter for SO aerosol is typically below 100 nm at 0.2% supersaturation, and likely closer to 80 nm for the aerosol population sampled during SOCRATES (Fossum et al., 2018; Mallet et al., 2025). This suggests that UHSAS100 may introduce an even greater underestimation of CCN02 compared to UHSAS100. To reflect the potential offset between UHSAS100 and CCN02, we conducted sensitivity tests by increasing the observed “CCN” by 8% and 40%, representing the lower and upper bounds of the CCN02:N100 ratio uncertainty, to examine how this affects our results (Section 3.3.3).

Having discussed uncertainty in the observations, we can turn our attention to uncertainty in the representation of processes in models. While the PPE samples a large number of possible representations of the underlying physics, it is still quite sparse (Lee et al., 2011). To systematically explore parametric uncertainty across the PPE, we build emulators (surrogate models) for the campaign-mean N_d and CCN using Gaussian Process (GP) regression (Watson-Parris et al., 2021). Emulators are trained by using the 45 perturbed parameters as inputs and simulation outputs (e.g., campaign-mean N_d and CCN) using a subset of the PPE ensemble as training data. Emulators are trained on the sample of different process representations in the CAM6 PPE data (Figure S2). The creation and validation of the emulators follows previous literature (Lee et al., 2011; Regayre et al., 2020; Song et al., 2024). With the GP emulators, we sample 1 million model realizations of N_d and CCN (e.g., model variants) with 1 million different combinations of parameter values sampled uniformly across 45 dimensional parameter space.

Model variants are ruled out when they are observationally implausible based on a implausibility measure

$$I(x) = \frac{|M - O|}{|Error(M)| + |Error(O)|} > 1 \quad (4)$$

205 where M is the emulator campaign-mean and O is the observed campaign-mean (Regayre et al., 2020). Error(M) and Error(O) denote the deviation from the emulator campaign-mean and observation campaign-mean, respectively. Error(M) comes from emulator uncertainty and the variance Var(M) in the emulator estimate is directly calculated from GP regression. Error(M) is estimated as $\pm 1.96 * \sqrt{Var(M)}$. The number of 1.96 is chosen as $\pm 1.96 * \sqrt{Var(M)}$ covers approximately 95% confidence bounds of the emulator uncertainty. Estimating observational uncertainty Error(O) as fractional value is commonly used in
210 observational constraints on models (Johnson et al., 2020; Song et al., 2024). We discuss observational uncertainty in terms of a fractional error f_{obs} . Finally, we write the implausibility metric I(x) where we account for 95% uncertainty in the emulator and an arbitrary observational uncertainty as

$$I(x) = \frac{|M - O|}{|\sqrt{Var(M)} \cdot 1.96| + |O \cdot f_{obs}|} > 1 \quad (5)$$

Model variants are excluded when I(x) exceeds 1. An illustration of our constraint process is summarized in Figure S3.

In this paper, we vary the observational uncertainty by varying f_{obs} under two conditions: (1) with emulator uncertainty and
215 (2) without emulator uncertainty, to characterize the impact of different sources of uncertainty on our ability to constrain the response of N_d to anthropogenic aerosol. We discuss the impact of different values of f_{obs} on the model constraint process in section 3.4. Eq 5 is a simplified implausibility metric as in Williamson et al. (2013); Johnson et al. (2020). Here, we only consider observational uncertainty and emulator uncertainty in the comparison between 1 million model variants with observations. Spatial-temporal representation uncertainty and model structural uncertainty are also important as discussed in
220 Johnson et al. (2020). We set the spatial-temporal representation uncertainty to 0 in Eq 5 as we collocated the model outputs to flight track locations in 2-min $\times$ 50-meter composites. The characterization of model structural uncertainty is conceptually ambiguous to quantify (Regayre et al., 2023) and is not considered in this work.

2.5 Constraint metric

We conducted sensitivity tests on the observationally plausible 2.5-97.5th percentile range of $\Delta N_{d,PD-PI}$ to the emulator
225 and observational uncertainties. The observationally plausible 2.5-97.5th percentile range of $\Delta N_{d,PD-PI}$ was calculated with varying presumed observation uncertainties. To reduce noise from imperfect emulators, we conduct another set of sensitivity tests with emulator uncertainties set to 0 (Error(M)=0) in Eq 4 in the sensitivity test. The 'constraint' is quantified as the reduction in the observational plausible range relative to the prior range predicted from the 1 million model variants. The relative reduction in range is calculated using

$$\text{constraint} = 1 - \frac{\text{Posterior: } \Delta N_{d,(PD-PI),\text{obs},97.5} - \Delta N_{d,(PD-PI),\text{obs},2.5}}{\text{Prior: } \Delta N_{d,(PD-PI),97.5} - \Delta N_{d,(PD-PI),2.5}} \quad (6)$$

230 Where the subscript 'obs' denotes the sources of observations that are used for constraints. The posterior range refers to the range of observationally plausible $\Delta N_{d,PD-PI}$ at 2.5-97.5th percentiles. The prior range refers the 2.5-97.5th percentile range

of $\Delta N_{d,PD-PI}$ derived from the original 1-million-member sample. Mathematically, the range of constraint in Eq 6 should vary between 0 to 1. The greater the magnitude, the better constraints we can achieve.

2.6 Spaceborne observation

235 In addition to the comparison between aircraft measurements and model outputs saved along flight tracks, we examine the simulated global oceanic mean N_d and confront it with observations. Observations of global oceanic N_d are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS is a passive radiometer onboard NASA's Terra and Aqua satellites. N_d is calculated from MODIS retrievals of effective radius (r_e) and optical depth (τ) assuming an adiabatic cloud (Grosvenor and Wood, 2014). MODIS N_d is calculated for daily means for the period 2003–2015 and is gridded to 1 by 1
240 degree resolution as in Grosvenor and Wood (2014). During winter, high-latitude regions (e.g., Arctic, Antarctic) have lower solar zenith angle (SZA), resulting in lower reflected solar radiation, making retrievals of cloud properties (e.g., r_e and τ) less reliable. MODIS N_d is unavailable during wintertime high latitude regions. To ensure consistency in the comparison between MODIS N_d and the model data, N_d data from months and latitudes where MODIS retrievals are unavailable are removed from the ESM dataset. In addition to global oceanic mean N_d from MODIS, we also examine a box region from MODIS with
245 latitude range of 65 S° - 42 S° and longitude range of 132 E° - 165 E°, which covers the SOCRATES campaign. MODIS N_d is computed in this box region and compared with campaign-mean N_d from SOCRATES in-situ measurements.

3 Results

As discussed above, previous studies have evaluated CAM6 in terms of its representation of SO aerosol, cloud, and precipitation characteristics using in-situ observations from SOCRATES (McCoy et al., 2021; Zhou et al., 2021; Gettelman et al., 2020;
250 McCluskey et al., 2023). They found that simulated N_d is typically too low in CAM6, which is similar to other ESMs (McCoy et al., 2020). However, finding why the SO N_d is low is complex since N_d is the result of sources and sinks (Wood et al., 2012; McCoy et al., 2020; Kang et al., 2022). To understand what leads to biases in N_d we need to simultaneously consider the impact of multiple processes to tackle the equifinality problem. Briefly, equifinality means multiple combinations of physical processes can result in the same observable state (i.e. N_d).

255 A common suggestion by previous studies is that investigation of aerosol, cloud, precipitation, glaciation, turbulence and activation processes is needed to understand the source of the model N_d bias (McCoy et al., 2021; Zhou et al., 2021; McCluskey et al., 2023). Here, we examine how different parameter combinations of these processes impact SOCRATES aerosol, clouds, and ACI in CAM6 in section 3.1. By observationally confronting simulations with different parameter combinations, we can evaluate the constrained parameter spaces (Section 3.2) and the likely range of observationally-plausible parameter spaces and
260 their associated range of $\Delta N_{d,PD-PI}$ (Section 3.3).

3.1 CCN, N_d and aerosol activation over SOCRATES in CAM6 PPE

We examine relationships between CCN and N_d across flight composites (50-meter $\times$ 2-min bin median) within individual PPE members. The number of flight composites valid for CCN- N_d comparisons from SOCRATES in-situ observations is 44 (Figure 2: red dots). This number is smaller than the results in McCoy et al. (2021) as we choose a lower altitude level for analysis with a focus on warm liquid cloud. The number of colocated flight composites valid for CCN- N_d comparisons for each PPE ensemble member (Figure 2: black dots) is less than that from observations (red dots) since some flight composites simulates near-zero N_d and are excluded from our analysis. This might be due to the coarse vertical resolution of CAM6 and linear interpolation cannot fully capture the N_d variability in the vertical. Despite the limitations, PPE ensemble members simulate CCN and N_d flight composites that are comparable with observations.

CCN at 0.2% supersaturation correlates positively with N_d when comparing matched flight composites along individual flight tracks in the PPE (Figure 2). This is not surprising as we expect CCN at 0.2% supersaturation to be a reasonable proxy for the aerosol particles that activate to form cloud droplets under typical marine boundary layer updraft conditions, consistent with observations (McCoy et al., 2021). Hereafter, we refer to the simulated CCN at 0.2% supersaturation from CAM6 simply as CCN for simplicity. Observations of CCN refer to the observed aerosol concentration with diameters ranging from 0.1 μm to 1 μm from UHSAS100.

Figure 2 shows a subsample of ensemble members with varying levels of agreement with observations, but a positive correlation between CCN and N_d in log space is found for most of the PPE members (i.e., 224 out of 262). However, the linear regression slope of N_d on CCN is high relative to observations for the majority of PPE members (Figure S4a). Because N_d is a product of both CCN activating into droplets and precipitation removing drops (Wood et al., 2012), a higher CCN- N_d slope in CAM6 does not necessarily indicate a higher simulated aerosol activation efficiency. This diagnostic is broadly telling us that more CCN is required in CAM6 PPE to produce the same amount of N_d through aerosol activation in the presence of coalescence scavenging compared to observations, particularly at low N_d concentration (e.g., Figure 2a). Lower N_d over SOCRATES is associated with increased precipitation rate and greater contribution of coalescence scavenging in controlling N_d (Kang et al., 2022). The negative correlation between N_d and precipitation rate is also found in the CAM6 PPE (Figure S5). The high bias in the regression slope of N_d on CCN in CAM6 PPE may indicate a stronger loss in N_d from overestimated coalescence scavenging at low N_d concentration in models. Additionally, the low-biased N_d may also be influenced by an underestimation of subgrid-scale vertical velocity, turbulence intensity, and other dynamical factors that suppress supersaturation and droplet activation. We verify our hypothesis in the discussion of parameter constraints of CAM6 PPE using observations in section 3.2.

Most of the PPE members simulate N_d that is low relative to observations, regardless of whether CCN is underestimated (Figure S4, Figure 2). One example is the PPE ensemble of 244 of CAM6 where even though the simulated CCN is relatively close to observations, the N_d is still biased low. (Figure 2d). This supports the hypothesis in McCoy et al. (2021) that aerosol biases are not the sole contributors to the low N_d in CAM6, highlighting the importance of other contributing factors.

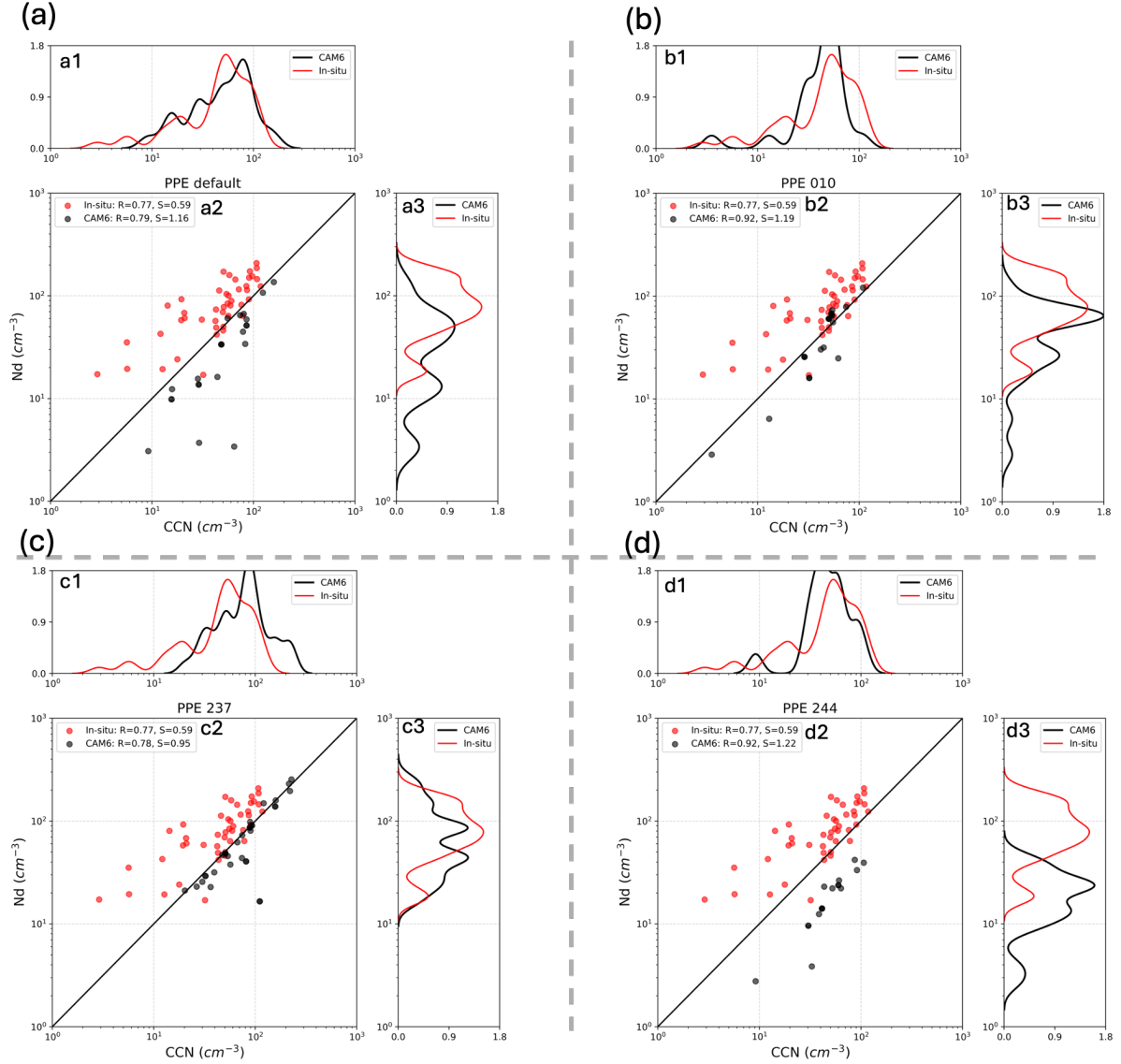


Figure 2. Relationships between SOCRATES CCN and in-cloud cloud droplet number concentration (N_d) from in-situ measurements (red) and CAM6 members (black), based on flight composites along individual flight tracks (scatters). Flight composite are constructed by binning observations into 50 m (altitude) by 2 min (time) bins for each flight. CAM6 PPE CCN and in-cloud N_d are collocated to observation composites (50 m $\times$ 2 min bins) by linear interpolation for individual PPE members. Bin medians are taken for comparison with CAM6 models following McCoy et al. (2021). CAM6 in-cloud N_d is computed as N_d divided by liquid cloud fraction (when cloud fraction $\leq 1\%$, we set $N_d = 0$). PDFs of number concentrations of CCN (top) and cloud droplets (right) for matched binned values occurring for CAM6 (black) and observations (red) are shown. (a) Default CAM6 configuration (i.e., PPE simulation for ensemble member 000), (b) PPE simulation for ensemble member 010, (c) PPE 237, (d) PPE 244. PPE members numbered 010, 237 and 244 are chosen to represent cases with varying levels of agreement between the simulated and observed CCN and N_d .

Although most PPE members exhibit low N_d , there are some members that are close to observations (Figure S4, Figure 2c).
 295 We next compare the PPE with observations to rule out PPE members that are far away from observations (e.g., Figure 2d) and
 characterize which parameterized processes are important to ACI over SO.

3.2 CAM6 parameter constraints from SOCRATES measurements

We compare campaign-means of CCN and N_d because averages reduce random errors of aircraft measurements due to instru-
 ment noise, atmospheric turbulence, or other transient variations (Schutgens et al., 2020). Unlike random errors, systematic
 300 errors such as sensor miscalibration and systematic sampling cannot be reduced by averaging. The model-observation compar-
 ison process follows Eq 5 as detailed in Section 2.4.

Observations of CCN and N_d identify and constrain physical processes that are important for ACI (Figure 3). We show the
 10 most constrained parameters out of 45 in Figure 3. The full list of constrained parameter spaces is shown in Figure S6. By
 examining how different parameter values are constrained relative to observables we can try to build an understanding of how
 305 different processes drive observables. This also illustrates the problem of equifinality where observed values can be arrived at
 by combining processes in different ways.

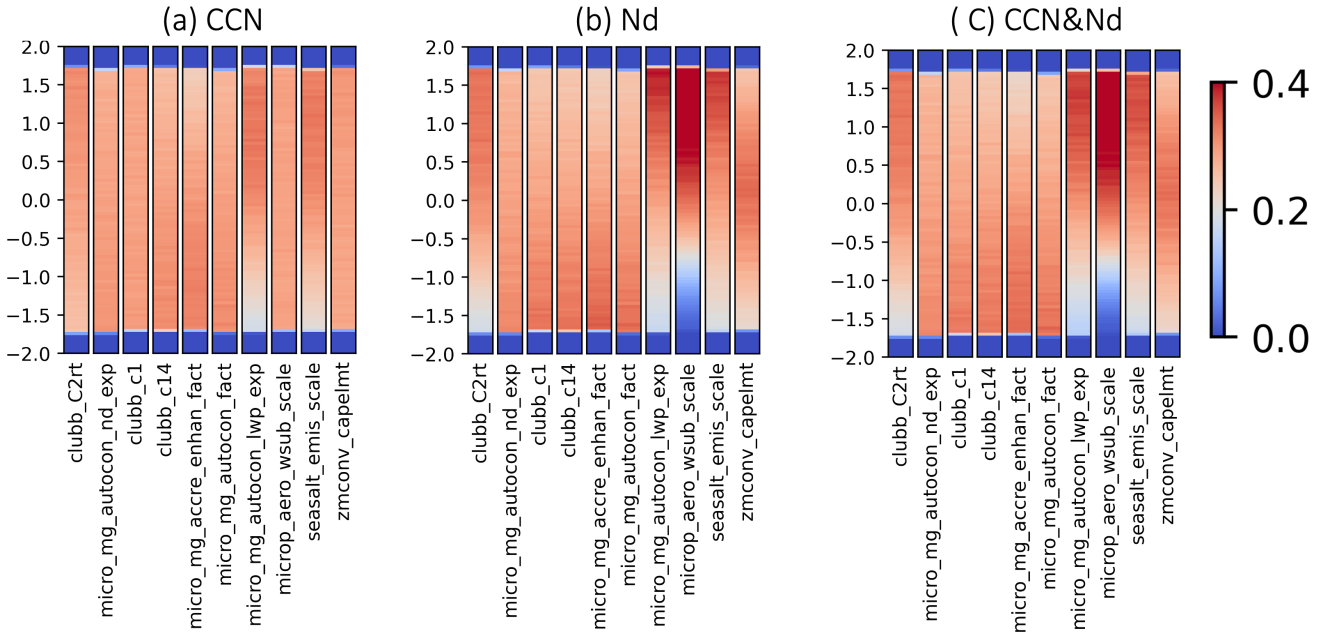


Figure 3. 10 parameters with constrained parameter spaces with observations of (a) CCN, (b) N_d and (c) CCN & N_d . Parameter spaces are standardized with mean 0 and variance 1. Warmer colors mean a higher intensity and more data points in that range.

Confronting the PPE with observations of CCN constrains aerosol processes (e.g. sea salt emission) and precipitation processes (e.g. autoconversion, accretion) (Figure 3a; the detailed parameter explanation is in Table S1). The sea salt emission

scale factor is constrained to higher values, indicating observations of CCN during SOCRATES are consistent with stronger
 310 aerosol production in the CAM6 PPE. This is consistent with a lack of aerosol production in CAM6 (McCoy et al., 2021; Zhou
 et al., 2021).

Constraints on precipitation processes point to the importance of precipitation as an aerosol sink. One of the key parameterization in warm cloud in climate models is the autoconversion, which represents the rate of initial rain formation through collision-coalescence between small cloud droplets.

315 We can make sense of the relationship between CCN and the autoconversion parameters by looking at how the rate of rain creation through autoconversion works in CAM6. Autoconversion in CAM6 is written as (Khairoutdinov and Kogan, 2000; Gettelman et al., 2015):

$$\left(\frac{\partial q_c}{\partial t}\right)_{\text{auto}} = a \cdot q_c^b \cdot N_d^{-c} \quad (7)$$

where $\left(\frac{\partial q_c}{\partial t}\right)_{\text{auto}}$ is the rate of generation of rain from cloud water. The autoconversion rate depends on the cloud droplet concentration (N_d [cm^{-3}] in Eq 7) and cloud water content (q_c [kg/kg] in Eq 7). a , b and $-c$ are uncertain parameters perturbed in
 320 the CAM6 PPE (Table S1). They are `micro_mg_autocon_fact`, `micro_mg_autocon_lwp_exp` and `micro_mg_autocon_nd_exp`, respectively. Selecting parts of parameter space that are consistent with observations of CCN leads to lower autoconversion scale factors (a in Eq 7) (less efficient rain production by cloud). The effect of larger exponents on liquid water content (b in Eq 7) on the rain production depends on the relative magnitude of liquid water content q_c . Larger b can result in thicker clouds that precipitate more efficiently under conditions of $q_c > 1$ kg/kg. A reversed effect can happen under conditions of
 325 $q_c < 1$ kg/kg. The condition of $q_c > 1$ kg/kg seems unlikely during SOCRATES campaign observations and model simulations (Khairoutdinov and Kogan, 2000; Gettelman et al., 2020). Overall, this results in a lower rain rate across cloud liquid water content values when the scale factor is minimized and the exponent is maximized for the liquid water content in typical stratocumulus clouds (Figure S7). The shift to lower rain rate with observational constraints in CAM6 PPE indicates that rain rate and the loss of CCN from precipitation scavenging are overestimated for the majority of members in CAM6 PPE.

330 Observations of N_d from SOCRATES constrain parameters related to aerosol and precipitation process (Figure 3b), consistent with the findings in Wood et al. (2012) and Kang et al. (2022); McCoy et al. (2020) that the N_d budget is a function of a source of droplets from CCN and sink from collision-coalescence. The constraint on initial rain formation rate during autoconversion (Eq 7) and the constraint on the strength of precipitation suppression are consistent with CCN observations (Figure 3ab). Broadly, constraints on precipitation formation are consistent with a weaker sink of cloud droplets as well as less cloud
 335 droplet removal via precipitation. This supports the hypothesis in Section 3.1 that rain rate and the loss of N_d from precipitation scavenging are overestimated for the majority of members in CAM6 PPE. We also want to note that the SO is dominated by supercooled liquid cloud (Gettelman et al., 2020; McCluskey et al., 2023), making the glaciation (Bergeron–Findeisen Process: water vapor deposits onto ice crystals) important in this region. This means that the growth of ice crystals might be an important sink for N_d . However, we believe the N_d loss from freezing is minimal when our analysis is restricted to the altitudes below 2
 340 km. This is because a large fraction of snow melts and contributes to rain precipitation at low altitudes (Figure 2 in Field and Heymsfield (2015)). Mixed-phase and ice cloud processes are important in initiating rain as most rain is derived from ice that

has melted to form rain (Bergeron, 1935). However, the importance of ice processes is not apparent in our process constraint focused on warmer clouds (Figure S6). McCluskey et al. (2023) examined ice processes over SOCRATES using observations of detailed aerosol and ice nucleating particle (INP) measurements and models (e.g., CAM6), but the process constraints on ice processes with observations of INPs is beyond the scope of this study.

The parameter constraints from observations of N_d are more stringent than the constraints resulting from using observations of CCN. This is consistent with N_d being the emergent product of aerosol and precipitation processes (Wood, 2012). In addition to aerosol and precipitation processes, mechanisms important for aerosol activation are also constrained by N_d observations (Figure 3b), such as deep convection (e.g., `zmconv_capelmt`), subgrid velocity (e.g., `microp_aero_wsub_scale`), and turbulence (e.g., CLUBB: Cloud Layers Unified by Binormals parameters in table S1), as they play a role in vertical aerosol transport and in generating supersaturation. In particular, `microp_aero_wsub_scale` is efficiently constrained to higher values, suggesting an underestimated subgrid velocity (i.e., lower updraft speed) that suppresses supersaturation, leading to lower N_d .

Finally, we examine the effect of constraining the PPE using observations of both N_d and CCN. The effect of combining these constraints is similar to the constraint arrived by N_d alone (Figure 3c). This is consistent with N_d being an emergent property of both aerosol processes and cloud and precipitation processes.

Observations of CCN and N_d during SOCRATES constrain aerosol, precipitation, and cloud processes. In the next section we examine whether the process constraint from observations of CCN and N_d constrains the response of N_d due to anthropogenic aerosol. Precipitation rate would be a useful constraint on the response of N_d as both observations of CCN and N_d constrain precipitation process as discussed in this section. However, the path to including an observational constraint of cloud base precipitation is somewhat opaque and is not included here. Light precipitation rate at cloud base can be retrieved using radar-lidar techniques (Kang et al., 2022), but to provide an apples-to-apples comparison to CAM6 in terms of cloud base precipitation we believe that an instrument simulator is needed (Silber et al., 2022).

3.3 Observationally plausible $\Delta N_{d,PD-PI}$ from SOCRATES measurements

As discussed in the previous section, observations of N_d and CCN constrain the range of possible process representations. In turn, these same processes drive the response of N_d to anthropogenic aerosol. This results in a strong correlation between PD N_d and PI N_d (Figure 4; black line) and by extension the change in N_d between PI and PD (Figure 4; orange line) in the CAM6 PPE. The fractional change in N_d ($\Delta \ln N_d$) is computed by taking the slope of $\Delta N_{d,PD-PI}$ to N_d following the definition from Bellouin et al. (2020). The $\Delta \ln N_d$ predicted by the CAM6 PPE (0.23) is greater than the expert elicitation range from Bellouin et al. (2020) (i.e., 0.05 to 0.17) (Figure 4). The emergent relationship between PD N_d (observable) and the $\Delta N_{d,PD-PI}$ (unobservable) with a r-value of 0.95 can be used to constrain the likely range of $\Delta N_{d,PD-PI}$ if we know the possible range of PD N_d (observable). As suggested in Klein and Hall (2015), emergent relationships used for constraints require process-level understanding. We explain the emergent behavior from CAM6 PPE (Figure 4) using a sink-source model of N_d in Section 3.3.2

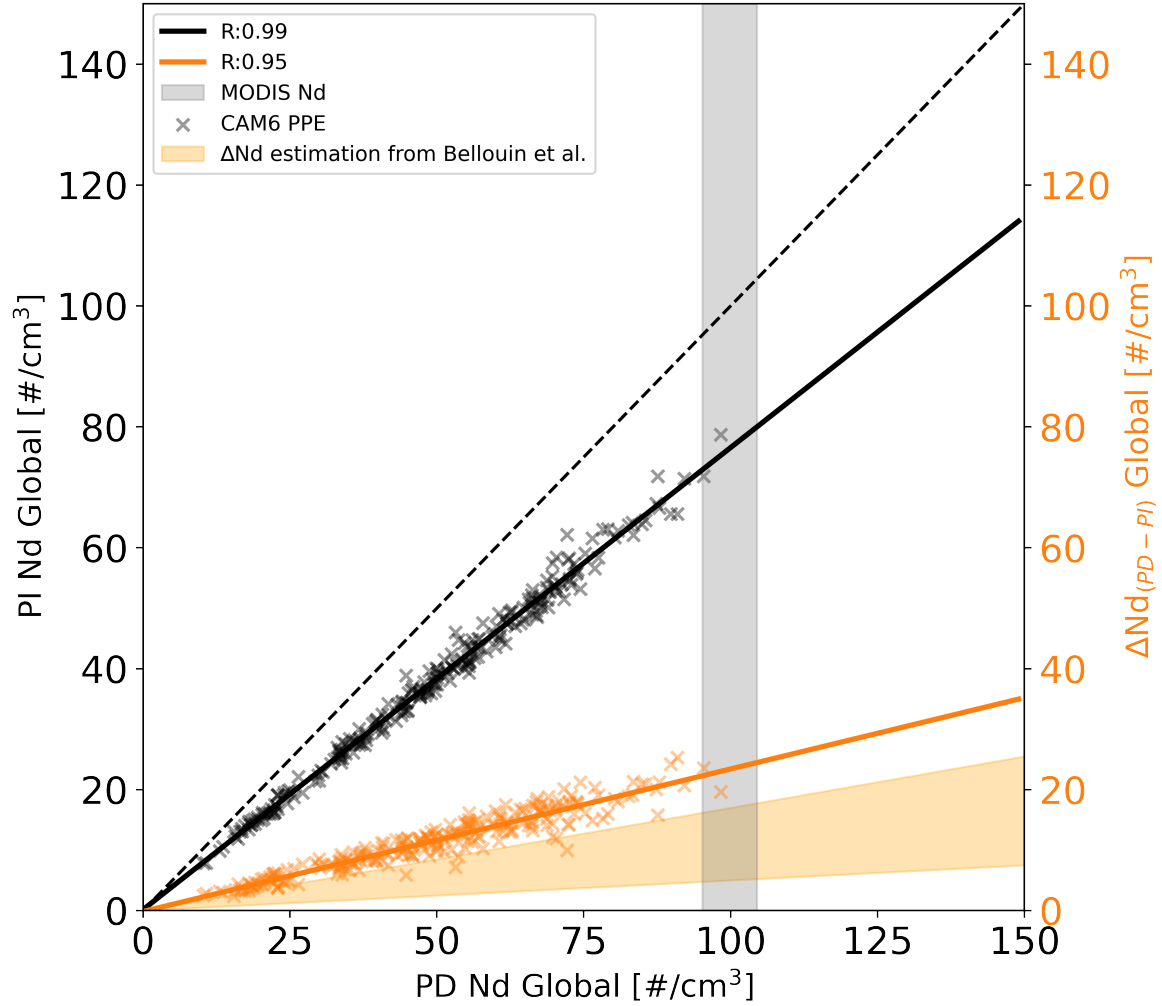


Figure 4. Global oceanic mean of preindustrial (PI) N_d (black) and $\Delta N_{d,PD-PI}$ (orange) as a function of present-day (PD) oceanic N_d from CAM6 PPE members (x-shaped markers). The 95% confidence on the interannual range of global oceanic mean N_d from MODIS is shown in the gray vertical bar. The estimated $\Delta N_{d,PD-PI}$ based on the fractional change in N_d ($\Delta \ln N_d$) from Bellouin et al. (2020) is shown in the orange shading.

3.3.1 Constraint using regional measurements

One question is whether SOCRATES, a field campaign over the SO where natural aerosols dominate, can be used to constrain the perturbation in N_d globally. SOCRATES samples natural aerosols and microphysical processes in a pristine environment (McCoy et al., 2021; McFarquhar et al., 2021), but it is not entirely isolated from the effects of anthropogenic aerosol emissions (Figure 1a). This is consistent with Hamilton et al. (2014), who show that the SO in the present day atmosphere is not always pristine. In addition to aerosol availability acting as a source for the N_d budget, both N_d and natural or anthropogenic aerosols share similar removal pathways through precipitation scavenging (Zheng et al., 2024; Wood et al., 2012; Kang et al., 2022), making the processes sampled during SOCRATES relevant for understanding the N_d perturbations on a global scale.

Another question is simply how representative is the N_d observed in the sample from SOCRATES of the global mean. Across the PPE, the campaign-mean N_d over SOCRATES correlates with the global, oceanic-mean of N_d with an explained variance of 0.36 across the PPE members (Figure 5). This positive correlation in N_d is reasonable as processes that govern droplet activation and removal of N_d share similarities over the global ocean and in the SO. The removal of N_d is primarily due to the precipitation scavenging (Wood et al., 2012; Kang et al., 2022). The relationship between the amount of CCN and the resultant N_d contain information about this sink term as well as the transport of CCN to cloud.

The relationship between SOCRATES N_d and the global mean also speaks to the importance of the marine, pristine baseline of aerosol in setting N_d . Previous studies underline the contribution of the oxidation of DMS (McCoy et al., 2015), sea spray (Wood et al., 2012; McCoy et al., 2015; Kang et al., 2022), and transportation of anthropogenic aerosols from continents (Wood et al., 2012; McCoy et al., 2018) to oceanic CCN.

Observational records also show consistency in the amount of N_d between the SO and the globe. Spaceborne observations of SOCRATES campaign-mean N_d and global-mean N_d are relatively consistent (Figure 5: black dot). In-situ campaign-mean N_d (Figure 5: gray dashed line) is slightly less than spaceborne observations (Figure 5: black dot), while the difference is small despite originating from entirely different methodologies.

We hasten to point out that we are not trying to argue that SOCRATES is sufficient to provide a complete picture of global-scale processes. However, SOCRATES does illustrate the utility of investigating even a single field campaign in this framework. Including additional campaigns in future field is likely to provide additional constraint on global-scale processes.

3.3.2 Constraint from present day observations

N_d sampled during SOCRATES contains information for globally-relevant processes (Figure 5), but do PD observations of aerosol and cloud properties constrain the anthropogenic perturbation in N_d ? We find this to be the case in the context of the PPE. To dissect the causes of the relationship between PD N_d and PI N_d , the relationship between PD N_d and $\Delta N_{d,PD-PI}$ and the relationship between CCN and N_d we turn to a simple budget model of N_d . Based on Wood et al. (2012); Kang et al. (2022), the N_d budget model is described as a function of source of CCN and sink from precipitation scavenging

$$N_{d,PD} = \frac{\left[CCN_{FT} + \frac{F(\sigma)U_{10}^{2.8}}{Dz_i} \right]}{\left[1 + \frac{hKP_{CB}}{Dz_i} \right]} \quad (8)$$

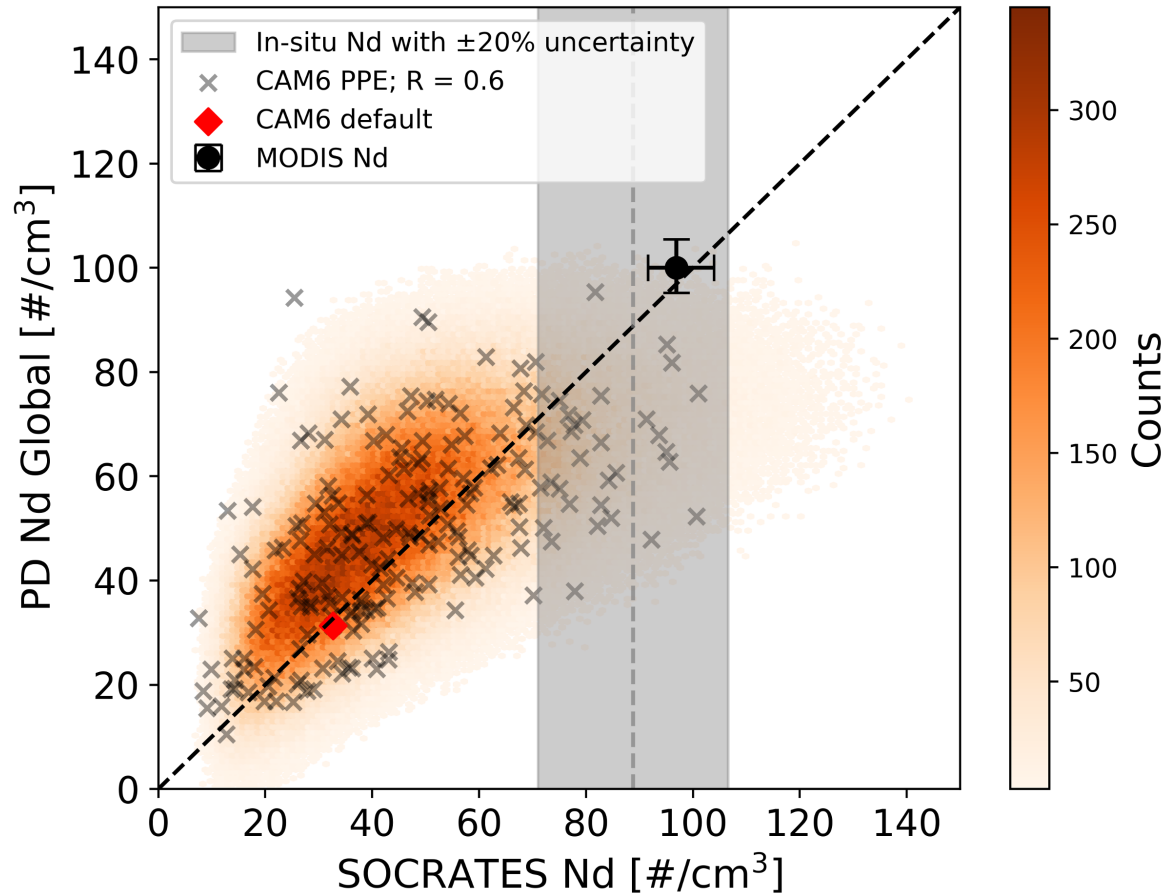


Figure 5. Global oceanic mean of present day (PD) N_d versus SOCRATES campaign-mean N_d from the CAM6 PPE members (x-shaped markers), 1M emulations from the PPE (orange color shading indicates density) and observations. The black dot shows the observational global-mean N_d and campaign-mean N_d calculated from MODIS, with its 95% confidence interval based on the interannual range. Observational SOCRATES campaign-mean N_d from SOCRATES in-situ measurements is shown as the vertical dashed line with an uncertainty of $\pm 20\%$ from the campaign-mean.

405 In Eq 8, the parameterized source of CCN is from free troposphere CCN_{FT} and surface contribution $\frac{F(\sigma)U_{10}^{2.8}}{Dz_i}$ (e.g., sea spray aerosol), where $F(\sigma)$ is the sea spray function that depends on supersaturation (Clarke et al., 2006) and Dz_i is the subsidence rate at cloud top, which is used as an approximate for entrainment rate (McCoy et al., 2020). For the precipitation sink term $\frac{hK P_{CB}}{Dz_i}$, h is the cloud thickness, K is a constant that depends on the collection efficiency of cloud droplets and P_{CB} is the rain rate at cloud base (Wood, 2006). The N_d budget model has been used to predict the N_d amount with confidence over the
 410 subtropic and mid-latitudes (Mohrmann et al., 2018; Zheng et al., 2018; Kang et al., 2022).

In this study, we follow the basic source-sink model idea from Wood et al. (2012) but we simplify the N_d budget model to fewer terms for a conceptual understanding of the relationships between variables. The N_d budget model is written as

$$N_d = \frac{\lambda \cdot CCN}{1 + k_{remo} \cdot P_{CB}} \quad (9)$$

Instead of parameterizing CCN source from free troposphere CCN_{FT} and surface contribution $\frac{F(\sigma)U_{10}^{2.8}}{Dz_i}$, we characterize CCN source as $\lambda \cdot CCN$, where λ is a scale factor that accounts for the amount of CCN that can be activated to cloud droplets
 415 depending on the vertical updraft, relative humidity, size and hygroscopicity of CCN, etc. λ varies from 0 to 1. For precipitation sink term $\frac{hK P_{CB}}{Dz_i}$, we simplified it as $k_{remo} \cdot P_{CB}$, where k_{remo} equals to $\frac{hK}{Dz_i}$. It accounts for the rate of loss of N_d from precipitation. To estimate k_{remo} , we set $K = 2.25 \text{ m}^2 \text{ kg}^{-1}$, subsidence rate $Dz_i = 4 \text{ mm s}^{-1}$ following Wood et al. (2012); Kang et al. (2022); McCoy et al. (2020). Cloud thickness h is set to 300 m, which is a typical magnitude for marine clouds (Wood, 2012). Changing cloud thickness h to smaller (e.g., 100 m) or larger values (e.g., 500 m) does not significantly change
 420 our results.

In this idealized set up, the CCN source ($\lambda \cdot CCN$) and precipitation sink ($k_{remo} \cdot P$) are set to be the same between PI and PD. This is a reasonable assumption for CAM6 PPE as its parameter setup is the same in the paired PI and PD simulations. The only difference between PI and PD is the amount of CCN. Therefore, we can write N_d budget model in PI and PD as

$$N_{d,PI} = \frac{\lambda \cdot CCN}{1 + k_{remo} \cdot P_{CB}} \quad (10)$$

$$N_{d,PD} = \frac{\lambda \cdot (CCN + \Delta CCN)}{1 + k_{remo} \cdot P_{CB}} \quad (11)$$

where ΔCCN in Eq 11 stands for CCN from anthropogenic aerosol emissions. Eq 10 and Eq 11 allow for a simplified
 425 representation of the underlying physical processes driving N_d budget in PI and PD. $\Delta N_{d,PD-PI}$ can be calculated by taking the difference between Eq 11 and Eq 10

$$\Delta N_{d,PD-PI} = \frac{\lambda \cdot \Delta CCN}{1 + k_{remo} \cdot P_{CB}} \quad (12)$$

We evaluate the relationship between $\Delta N_{d,PD-PI}$ and PD N_d , CCN and N_d using the budget models. We calculate the $\Delta N_{d,PD-PI}$ and PD N_d in response to an anthropogenic perturbation of $+\Delta CCN$ at varying precipitation rate P_{CB} and CCN scale factor λ . With this simplified set up, $\Delta N_{d,PD-PI}$ and PD N_d , CCN and N_d are positively correlated (Figure 6),
 430 consistent with the CAM6 PPE (Figure 2, Figure 4). At larger precipitation rate P , more CCN is needed to activate to form the

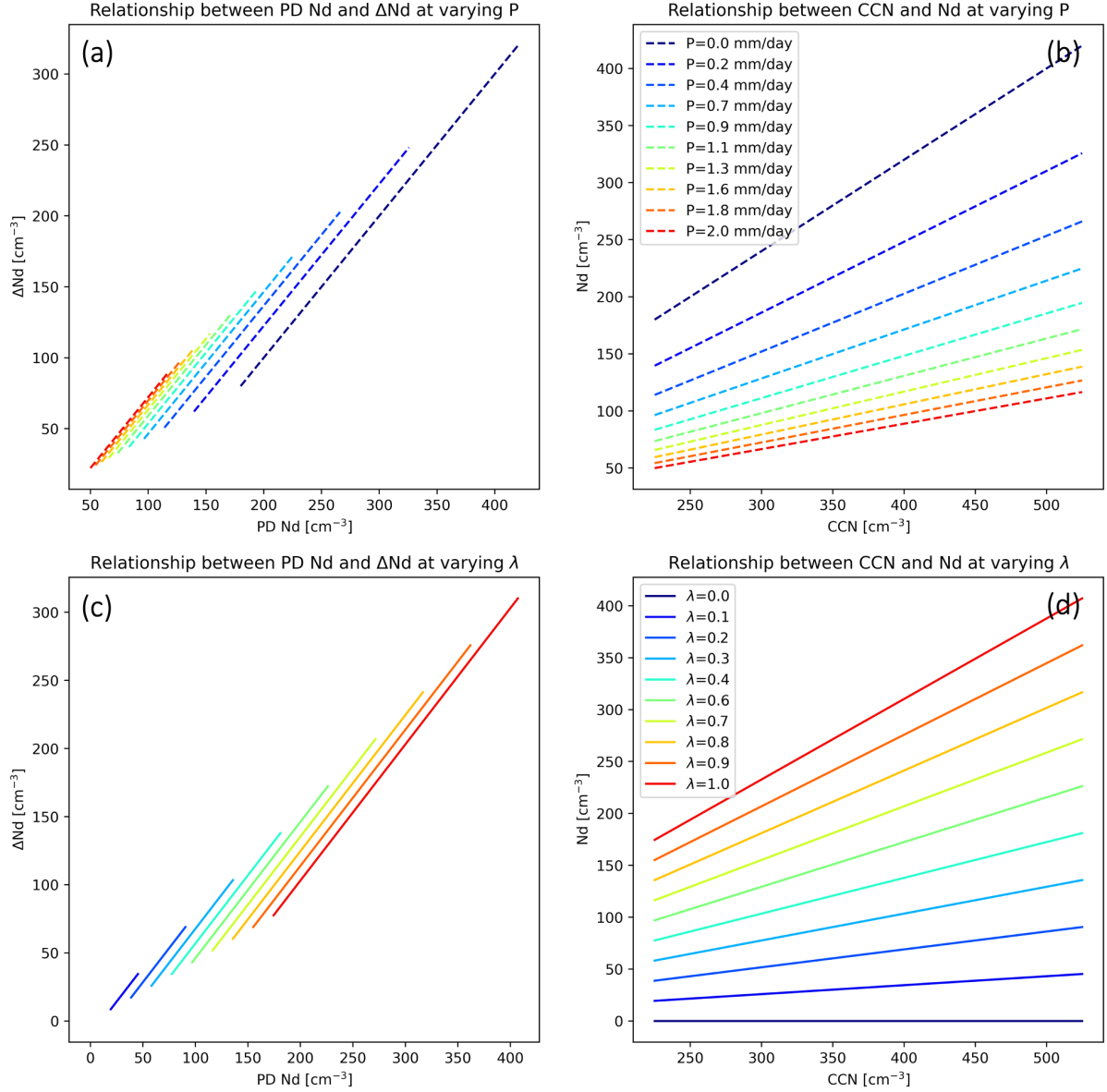


Figure 6. Idealized relationships based on the source-sink model of the N_d budget (Eq 8 in Wood et al. (2012) with modifications). (a) $\Delta N_{d,PD-PI}$ versus PD N_d based on Eq 11 and Eq 12 at varying precipitation rate P . P is set to vary from 0 to 2 mm/day in 10 equal increments. The varying P is within the observational range in Wood et al. (2012). CCN is set to 125 cm^{-3} as a background CCN from natural source. ΔCCN is set to be varying between 100 to 400 cm^{-3} with 20 equal increments. k_{remo} is set to 0.8. (b) N_d versus CCN at varying precipitation rate P with the same model setup as (a). (c) $\Delta N_{d,PD-PI}$ versus PD N_d at varying CCN scale factor λ . Precipitation rate P is set to 0.2 mm/day. (d) N_d versus CCN at varying λ with the same model setup as (c).

same amount of N_d (Figure 6b). This leads to an overall lower N_d and N_d change due to aerosols at high precipitation rate (Figure 6a). With more CCN amount (or larger λ), there is a higher N_d and N_d change due to aerosols (Figure 6cd). These results suggest that the positive correlations in CAM6 PPE members are driven by sink from precipitation scavenging and source from CCN as depicted in the idealized model.

435 Understanding the positive relationships in CAM6 PPE we can use PD observations of N_d to constrain unobservable quantities such as perturbations in N_d due to anthropogenic aerosols. To constrain global-mean quantities, we use observations of N_d from SOCRATES campaign as the variance in global oceanic-mean N_d is largely explained by SOCRATES campaign-mean N_d (Figure 5). In addition to PD N_d observations, we use the observed aerosol concentration from UHSAS100 from SOCRATES as a proxy of CCN to provide an additional constraints on $\Delta N_{d,PD-PI}$ as the parameter spaces of the PPE have
 440 been shown to be constrained by CCN in Section 3.2 (Figure 3a). Precipitation processes are constrained by both observations of CCN and N_d as discussed in section 3.2. Inspired by this, we wanted to examine the effects of cloud base precipitation on the constraints on $\Delta N_{d,PD-PI}$. However, we found it difficult to make a direct comparison between CAM6 and cloud radar-lidar-retrieved precipitation rates at cloud base. Therefore, our constraints on $\Delta N_{d,PD-PI}$ focus on observations of CCN and N_d . Nonetheless, we provide an illustration of what the constraints would behave if observed precipitation rates were
 445 used, based on idealized sensitivity tests discussed in Section 3.3.3.

3.3.3 Constraints from CCN and N_d measurements

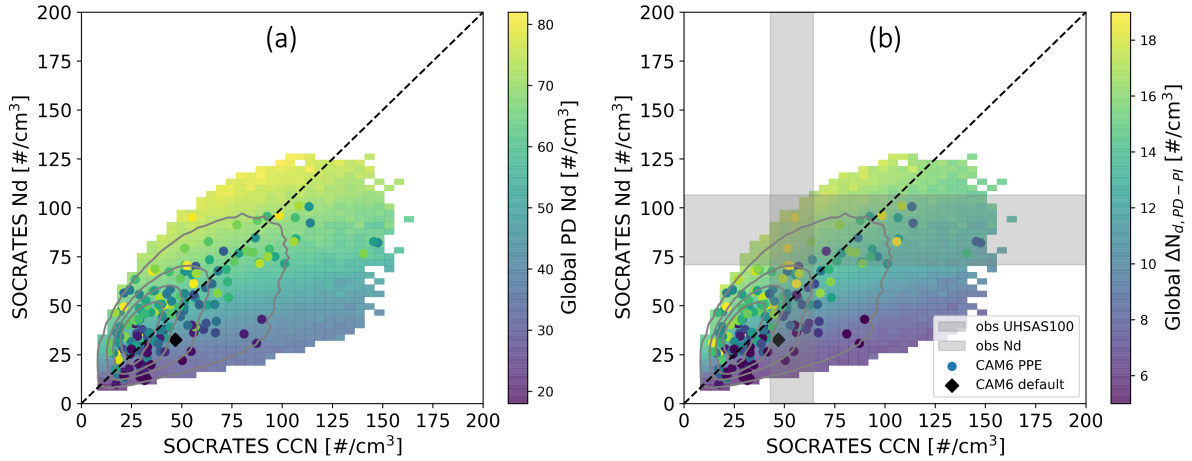


Figure 7. (a) SOCRATES campaign-mean N_d versus campaign-mean CCN and colored by present-day N_d from the CAM6 PPE members (color dots) and 1M emulations from the PPE (color shading). Emulate density is shown in solid contours. (b) The same with (a) but colored by $\Delta N_{d,PD-PI}$. The color shading shows 2D bin-averaged values of (a) global mean N_d and (b) $\Delta N_{d,PD-PI}$, computed using 60×60 bins in SOCRATES CCN and SOCRATES N_d space. This smoothing highlights large-scale patterns while excluding sparsely sampled regions. Colored points show individual PPE members without averaging. Observational SOCRATES campaign-mean CCN (i.e., UHSAS100) and N_d from SOCRATES in-situ measurements is shown as the gray shaded bars with an uncertainty of $\pm 20\%$ from the campaign-mean.

Before going into the constraints, we first examine the aerosol activation across the PPE members with different parameterizations. Campaign-mean CCN and N_d are positively correlated across PPE ensembles (Figure 7a), consistent with the CCN- N_d relationship across flight composite in individual models (Figure 2) and the idealized model (Figure 6ac).

450 Understanding emergent relationships is essential in constraining unobservable quantities (Klein and Hall, 2015). With the idealized N_d budget model (Figure 6), we have understood how the physical processes (i.e., source of N_d from CCN; sink of N_d from precipitation scavenging) are related to the correlations between variables across the CAM6 PPE (Figure 2, Figure 4). Next, we use airborne observations of CCN and N_d to rule out implausible model variants out of the 1 million variants emulated from the CAM6 PPE (Figure 7b) following the implausibility metric (Eq 5). The prior 2.5-97.5th percentile range of
 455 $\Delta N_{d,PD-PI}$ is 3.6 cm^{-3} to 19.8 cm^{-3} . Based on the implausibility metric (Eq 5), observations of CCN have no effect on the constraints on $\Delta N_{d,PD-PI}$. Using the same implausibility metric, observations of N_d over SOCRATES constrains the range of $\Delta N_{d,PD-PI}$ to be 6.1 cm^{-3} to 20.4 cm^{-3} at the 2.5-97.5th percentile, equivalent to 12% reduction in range and the median increases by 16% (Figure 8a). The constraints on $\Delta N_{d,PD-PI}$ are also consistent with the results in Song et al. (2024) that utilizes hemispheric contrast of N_d as a proxy to $\Delta N_{d,PD-PI}$ using observations of remote sensing from MODIS, indicating the consistencies between measurements of N_d from different observing techniques.

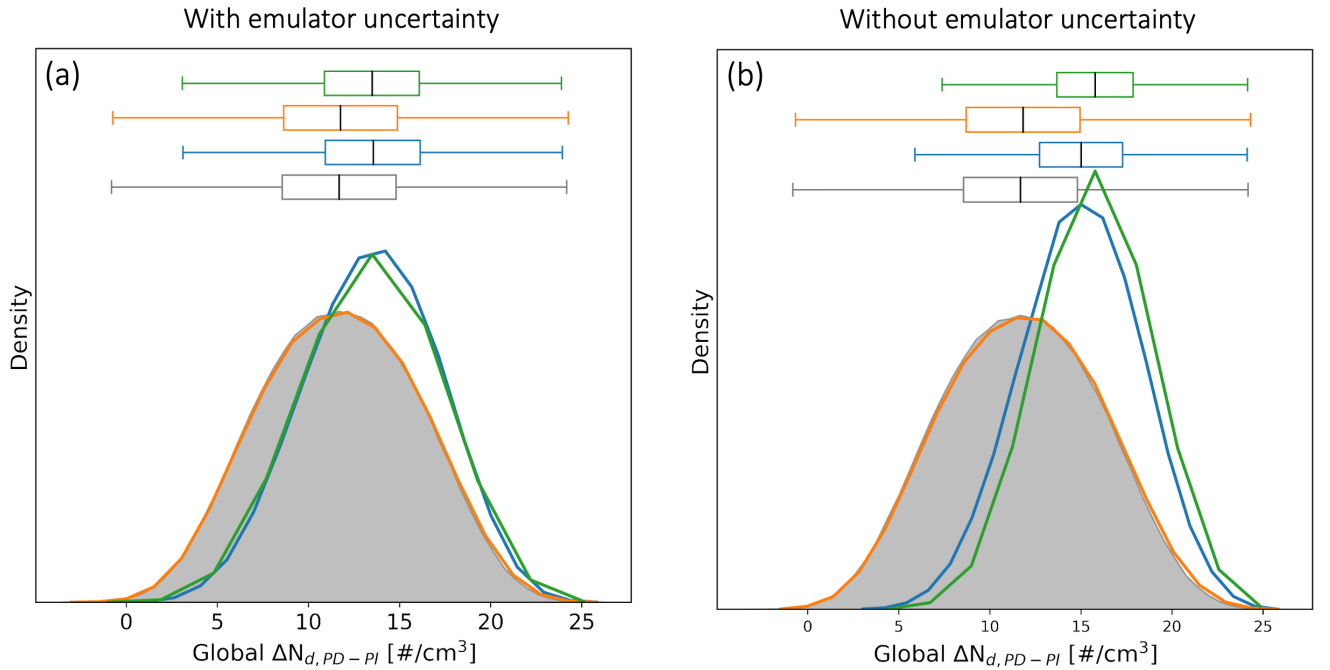


Figure 8. The distribution of emulated $\Delta N_{d,PD-PI}$ prior (grey shading), and observationally-constrained posterior from SOCRATES observed CCN only (orange), N_d only (blue), and CCN and N_d (green). (a) With emulator uncertainty. (b) without emulator uncertainty.

In this study, the emulator predictions are based on emulator mean predictions (i.e., M in Eq 4) and emulator uncertainties (i.e., $\text{Error}(M)$ in Eq 4). Although the emulator mean predictions are overall good, the emulator uncertainty created for CAM6 PPE outputs are relatively large (Figure S2). This may have a huge impact on the model-observation comparison process (Figure S3, Eq 4). We therefore examine the constraints on $\Delta N_{d,PD-PI}$ without the effects of emulator uncertainties (i.e., set
 465 $\text{Error}(M)=0$ by setting the number of variance to 0 (i.e., $0 * \sqrt{\text{Var}(M)}$) in Eq 5) to check whether this change significantly affects our constraints. The implausibility metric in this case follows

$$I(x) = \frac{|M - O|}{|O * 20\%|} > 1 \quad (13)$$

Consistencies are found in the observational constraints on $\Delta N_{d,PD-PI}$ under conditions both with and without emulator uncertainties. Observations of CCN have no effect on the constraints under both conditions (Figure 8), suggesting that the zero constraint from CCN is not a result of large emulator uncertainties. Instead, the near zero constraint on $\Delta N_{d,PD-PI}$
 470 from CCN might be because the CCN provides less information about the number of cloud droplets that can form through aerosol activation compared to direct measurements of cloud droplet numbers. Although CCN number concentration at a given supersaturation accounts for a certain level of chemical composition of aerosols (i.e., hygroscopicity and size), environmental conditions, which is critical for their activation to cloud droplets, is less known. The observational constraints on $\Delta N_{d,PD-PI}$ from SOCRATES N_d are consistent in both conditions in terms of the positive shift in the likely range of $\Delta N_{d,PD-PI}$.
 475 Observations of N_d narrow the $\Delta N_{d,PD-PI}$ range more efficiently under the condition without emulator uncertainties than under the condition with emulator uncertainties. The reduction in range of $\Delta N_{d,PD-PI}$ is 21% and the increase in the median is 28% when calculated without emulator uncertainties (Figure 8b).

Observations of CCN and N_d consistently constrain aerosol and precipitation processes as we discussed in section 3.2 so we examined their joint effects on $\Delta N_{d,PD-PI}$ in Figure 8. We found great improvement on the N_d constraint when including
 480 the effects of CCN, assuming no emulator uncertainty (Figure 8b). The $\Delta N_{d,PD-PI}$ range is narrowed down by 27% and the median shifts from 11.7 cm^{-3} to 15.5 cm^{-3} (i.e., 35% increase in median). The result suggests that the direction (e.g., positive or negative shift) of the constraint on $\Delta N_{d,PD-PI}$ is not sensitive to emulator uncertainty, while the strength of constraints is sensitive to emulator uncertainty.

Precipitation scavenging works as a sink for both CCN and N_d (Figure 6ab), suggesting the strong potential of using pre-
 485 cipitation rate as an observational constraints on $\Delta N_{d,PD-PI}$. We found it is difficult to make apples-to-apples comparison between CAM6 and cloud radar-lidar retrieved precipitation rate at cloud base. Therefore, we do not use observed precipitation rates in this study. Instead, we examine what the constraints on $\Delta N_{d,PD-PI}$ would respond under two hypothetical campaign-mean surface precipitation rate constraints, used as idealized sensitivity tests. The results suggest that surface precipitation rate has no constraint on $\Delta N_{d,PD-PI}$ (Figure S8). The zero constraint might be due to surface precipitation rate is less informative
 490 than cloud base precipitation rate on the N_d budget due to the evaporation during descent. Another possible explanation is that while the precipitation sink explains a lot of variance from flight to flight, (Kang et al., 2022), it doesn't vary as dramatically between ensemble member representations of the entirety of the campaign mean because it is strongly controlled by the amount of water vapor and circulation. Overall, we believe further development of instrument simulators for simulating cloud

base precipitation rate for global models is needed to improve our ability to leverage airborne cloud base precipitation rate to
495 constrain global behavior.

As discussed in Section 2.4, using UHSAS100 as a proxy to CCN at 0.2% (CCN02) supersaturation may underestimate the observed CCN02. We conduct a sensitivity test on the constraints on $\Delta N_{d,PD-PI}$ by increasing observed CCN by 8% to 40%, according to the CCN02:UHSAS100 ratio uncertainty shown in Figure S2a in McCoy et al. (2021). The results suggest that increasing the observed CCN does not significantly affect the constraint on $\Delta N_{d,PD-PI}$ (Figure S9).

500 3.4 Sensitivity tests on the observationally plausible ΔN_d

The uncertainty associated with airborne measurements of aerosol and cloud microphysics is difficult to define as a fixed value because it depends on multiple sources of uncertainties such as: sampling error due to limited spatial and temporal coverage of flight tracks, variability in flight patterns (e.g., altitude, and positioning relative to cloud features), instrument noise due to environmental variability (e.g., turbulence, wind shear). Therefore, assuming a fixed observation uncertainty of $\pm 20\%$ for
505 airborne measurements in Section 3.3.3 is just a to provide a baseline amount of uncertainty.

In this section, we perform sensitivity tests on the observationally plausible $\Delta N_{d,PD-PI}$ by the varying observational uncertainty under conditions of with emulator uncertainty (i.e., $\text{Error}(M) = \pm 1.96 * \sqrt{\text{Var}(M)}$ in Eq 5) and without emulator uncertainty (i.e., $\text{Error}(M) = 0 * \sqrt{\text{Var}(M)}$ in Eq 5). Systematic uncertainty in the observations is assumed to vary from $\pm 5\%$ to $\pm 100\%$ for the campaign-mean. The constraints on $\Delta N_{d,PD-PI}$ is calculated as the reduction in the 2.5-97.5 percentile
510 range of $\Delta N_{d,PD-PI}$ following Eq 6.

Figure 9 shows the constraints on $\Delta N_{d,PD-PI}$ by observations of CCN (orange line), N_d (blue line) and combining CCN and N_d (green line). Overall, $\Delta N_{d,PD-PI}$ is more tightly constrained without emulator uncertainty (Figure 9bd), which makes sense as excluding emulator uncertainties (i.e., Eq 13) allows more model variants to be excluded during model-observation comparison relative to Eq 5. The constraints on $\Delta N_{d,PD-PI}$ under two conditions (i.e., with and without emulator uncertainty)
515 both have a positive shift in the 2.5-97.5 percentile range (Figure 9cd), suggesting the constraints are not a result of noise from emulator uncertainties.

Consistent with Figure 7, observations of CCN provide nearly no constraint on $\Delta N_{d,PD-PI}$ regardless of the uncertainty in CCN (Figure 9: orange line). With increasing uncertainties from airborne measurements of N_d , the plausible range of $\Delta N_{d,PD-PI}$ is less constrained under both conditions of with and without emulator uncertainty (Figure 9: blue line). The
520 weakened constraints on $\Delta N_{d,PD-PI}$ with increasing uncertainties in observations is due to the retention of more model variants as plausible during model-observation comparison process (Figure S3), thereby exacerbating the equifinality problem (Johnson et al., 2020), for which increasing number of plausible parameter combinations results in broader parameter spaces and thereby reduced constraints on $\Delta N_{d,PD-PI}$ (Figure 9ab).

The reduction in ranges (constraints) becomes negligible with observation uncertainty of N_d reaching $\pm 75\%$ under the
525 condition with emulator uncertainty (Figure 9a). The threshold of observational uncertainty of N_d that can achieve constraints is a bit larger under the condition without emulator uncertainty, which is $\pm 85\%$ (Figure 9b), suggesting skillful emulators play an important role in advancing our constraints. Again, we view uncertainty from emulation as an eminently tractable problem

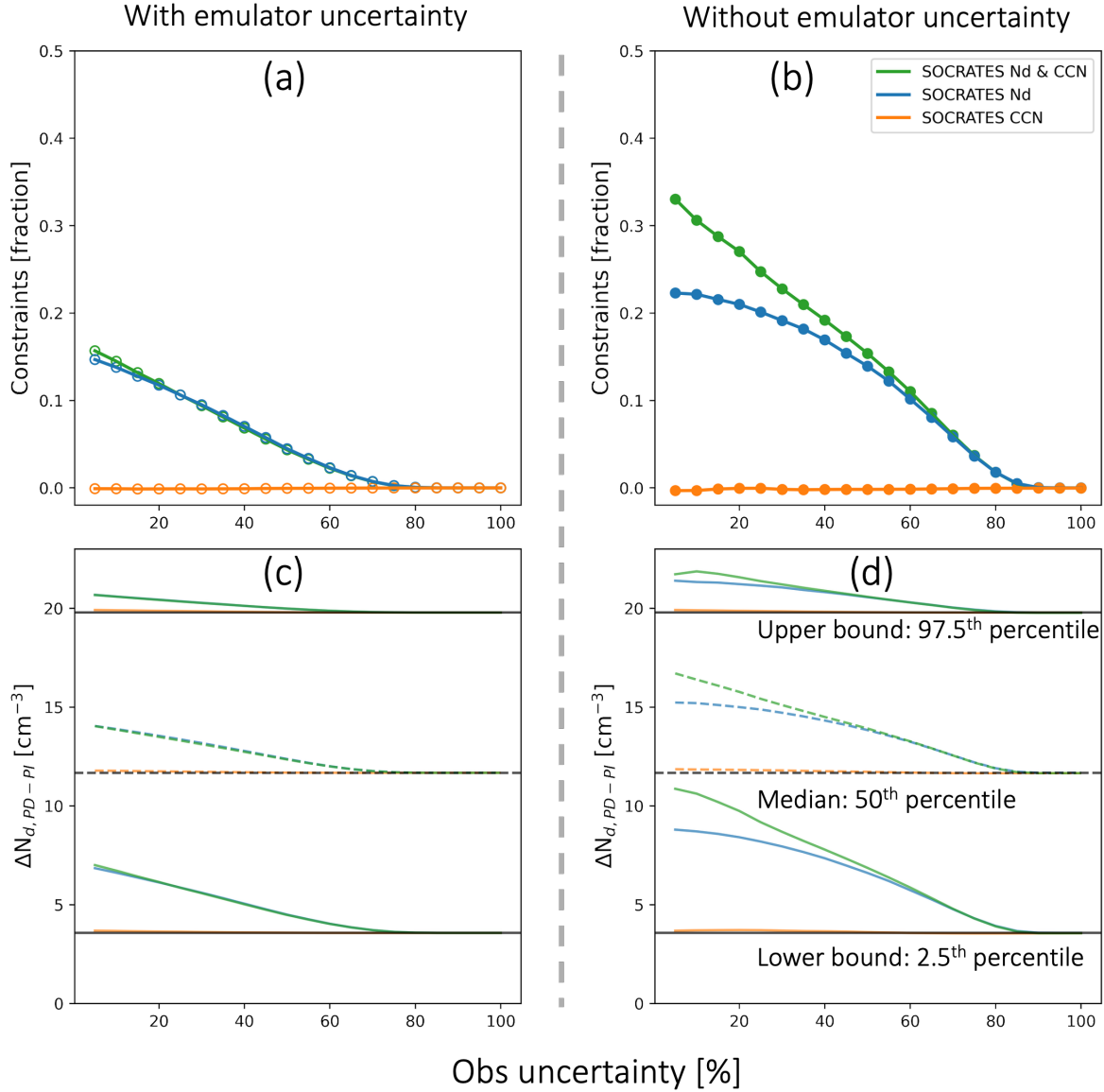


Figure 9. (a) The reduction in the observationally-constrained 2.5-97.5 percentile range of $\Delta N_{d,PD-PI}$ relative to the emulation prior range with increasing aircraft measurements uncertainties with emulator uncertainties considered, following the implausibility metric as described in Eq 5. Observations are from SOCRATES CCN from UHSAS100 (orange), N_d from CDP (blue) and both observations (green). Uncertainty ranges from $\pm 5\%$ to $\pm 100\%$ from the campaign-mean with 5% equal increments. (b) The same (a) but without emulator uncertainty, following the implausibility metric described as Eq 13. (c) The 2.5th, 50th and 97.5th percentile value of the observationally-constrained $\Delta N_{d,PD-PI}$ from SOCRATES CCN (orange), N_d (blue) and both lines of observations (green) as a function of aircraft measurements uncertainties. (d) The same with (c) but without the effects of emulator uncertainty.

compared to developing a better understanding of instrumental systematic uncertainty. Improving emulator uncertainty can be achieved by creating PPEs that can be easily emulated through a more careful selection of perturbed parameters and an increased amount of training data.

Another key insight from Figure 9 is that the improvement in constraint when both N_d and CCN are considered emerges at low observational uncertainties for both measurements (Figure 9ab: green line is above the blue). The improved constraint is also found in Regayre et al. (2023) that adding the number of constraint variables with no structural inadequacies improve the constraints on aerosol forcing when their constraints are consistent across multiple observation types. In our work, although the constraints on $\Delta N_{d,PD-PI}$ by observations of CCN is minimal, the constraints on parameter spaces by airborne observations of CCN and N_d are consistent (Figure 3), resulting in the improved constraints on $\Delta N_{d,PD-PI}$. The improvement of the constraints on $\Delta N_{d,PD-PI}$ only happens with small uncertainties associated with airborne measurements (Figure 9ab: green line), highlights the importance of accurate airborne measurements of aerosol and cloud properties on constraining $\Delta N_{d,PD-PI}$.

4 Conclusions

We use observations of aerosol (i.e. CCN) and cloud properties (i.e. N_d) from airborne in-situ measurements taken during SOCRATES (the Southern Ocean Clouds, Radiation, Aerosol Transport Study, Figure 1) to constrain model parameters related to aerosol-cloud interactions. To systematically examine constraints on parameterized processes, we used a PPE that varies 45 parameters related to aerosol-cloud interactions. While a large number of ensemble members (i.e., 262) were integrated across 45-dimensional parameter space, this sampling was still very sparse in an absolute sense. To better map this space, we trained statistical emulators (Figure S2) to create a set of 1 million model variants. Each model variant was compared against observations and are retained if its implausibility is less than 1 based on the implausibility metric (Figure S3, Eq 4). Our constraint on processes in this framework (Figure 2, 3) resulted in a constraint on the anthropogenic perturbation to N_d ($\Delta N_{d,PD-PI}$) (Figure 8).

Airborne observations of CCN and N_d over SOCRATES both constrain parameter spaces related to aerosol emission and precipitation processes (Figure 3), providing insights on developing in-situ instruments targeting these processes to better constrain cloud properties. Observations of N_d are more effective than CCN at constraining parameter space (Figure 3ab).

With constrained parameter space with observations of CCN and N_d (Section 3.2), we examine the likely range of $\Delta N_{d,PD-PI}$ at the constrained parameter space. One key result is that observations of CCN alone have minimal constraint on $\Delta N_{d,PD-PI}$, but the constraint from observations of N_d is strong (Figure 8). This is sensible because aerosol concentrations from UH-SAS100 give information about the aerosol population, but they do not directly inform the activation of aerosols into droplets.

To explain why the constraints work when using observations of N_d , we use idealized sink-source models of N_d (Eq 11, 10). Previous work identifies the precipitation sink of droplets as a key driver of N_d variability (Kang et al., 2022; Wood et al., 2012; McCoy et al., 2020). We extended these budget models to anthropogenic perturbations to N_d (Figure 6), and used the result to understand relationships emergent from the CAM6 PPE (Figure 2, 4, 7). Within the N_d budget model framework, we

560 identified both the precipitation sink and how CCN form droplets as important in controlling the positive correlation between CCN and N_d (Figure 6bd), the positive correlation between present day (PD) N_d and $\Delta N_{d,PD-PI}$ in CAM6 PPE (Figure 6ac).

Understanding the emergent behaviors from the CAM6 PPE (Figure 4), we next note that the amount of campaign-mean N_d over SOCRATES is a good approximate of the global oceanic-mean N_d (Figure 5). Within this framework, observations of N_d over SOCRATES constrains the amount of global $\Delta N_{d,PD-PI}$ due to anthropogenic aerosol influence in the CAM6 PPE, 565 with the strong positive correlation between PD N_d and $\Delta N_{d,PD-PI}$ in CAM6 PPE (Figure 4), which can be explained in the context of N_d sources and sinks (Figure 6).

We find that the range of parameters associated with precipitation processes that are consistent with observations of CCN and N_d is narrow, suggesting the potential of using precipitation rate as an observational constraints on $\Delta N_{d,PD-PI}$. We examine what the constraints would be like if we know the observed surface precipitation. However, the constraints from two 570 hypothetical surface precipitation is minimal (Figure S8). We did not include cloud base precipitation rate as a constraint due to the difficulty in comparing the simulated cloud base precipitation from CAM6 PPE with cloud base precipitation retrieved from cloud radar and lidar from SOCRATES. We identify further development of instrument simulators for global models as a useful avenue to improve our ability to leverage airborne data (e.g., cloud base precipitation) to constrain global behavior.

We find the constraints are sensitive to the implausibility setup. Overall, observations of N_d constrain $\Delta N_{d,PD-PI}$ to a 575 12% reduction in range and a 16% increase in the median, under the condition of with emulator uncertainty (Figure 8a). The constraint is improved under condition of no emulator uncertainty, for which the reduction in range of $\Delta N_{d,PD-PI}$ is 21% and the increase in the median is 28% (Figure 8b). The results suggest reducing emulator uncertainty is important for improving our constraint. More skillful emulators can be achieved by a careful selection of parameters focusing on key processes affecting ACI, particularly aerosol activation, cloud microphysics, and precipitation process as we discussed earlier in Section 3.2. 580 Additionally, this work provides insights in the possible parameters ranges in creating PPEs in the discussion of parameter constraints in Section 3.

We examine the sensitivity of the constraints on $\Delta N_{d,PD-PI}$ with observational uncertainties (Figure 9). We find discarding parameter combinations that don't mesh with observed CCN and N_d during SOCRATES yields an improved constraint on $\Delta N_{d,PD-PI}$ when we disregard emulator uncertainty and when observational uncertainty decreases (Figure 9ab). This 585 highlights the importance of considering systematic uncertainties in observations and continuing to develop our understanding of systematic uncertainty in observations of microphysics as well as designing campaigns that allow for stochastic sampling and more direct comparisons to ESMs.

In future work, incorporating additional in-situ constraints, such as aerosol composition, size distributions, or lidar-radar-retrieved cloud and precipitation properties could further narrow the range of plausible PPE configurations. Alternatively, 590 adding more variables as observational constraints may expose structural model uncertainties if the observations are incompatible with any members of the PPE ensemble (Regayre et al., 2023). Instrument simulators that translate model outputs into observation-like quantities (e.g., cloud-base precipitation rate) will also be essential for consistent comparisons. Moreover, incorporating variables such as ice-nucleating particles or ice crystal concentrations, could extend the PPE framework to mixed-phase regimes. While such extensions are beyond the current scope of this study, which focuses on warm liquid clouds,

595 they represent promising directions for future work. Together, these directions can improve the use of PPEs in constraining aerosol–cloud interactions.

Code and data availability. The description of the PPE dataset is in Eidhammer et al. (2024). The PPE dataset used in this study is stored on the NCAR-Wyoming Super-computing Center for community access. The aerosol and cloud measurements from SOCRATES are available at <https://data.eol.ucar.edu/dataset/552.002>. Cloud droplet number concentration from MODIS is available at online in NetCDF format from
600 the Centre for European Data Analysis (CEDA) (<https://catalogue.ceda.ac.uk/uuid/cf97ccc802d348ec8a3b6f2995dfbfff/>).

Author contributions. CS conducted the simulations of the PPE with configurations specific to this study with the help from TE and AG, conducted analysis of the data and wrote the manuscript. DTM initiated the idea of producing PPE simulations integrated along flight tracks, conducted analysis of the data and assisted in writing the manuscript. ILM contributed to the analysis of campaign observations and PPE outputs, provided input on the work and edited the manuscript. HB provided input on the manuscript and assisted in writing the manuscript.
605 AG and TE assisted in running scripts for the PPE simulations and assisted in editing the manuscript. DB assisted in editing the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. CS, DB, and DTM were supported by NASA Grant 80NSSC21K2014 and DTM was supported by the U.S. Department of Energy’s Atmospheric System Research Federal Award DE-SC002227; U.S. Department of Energy’s Established Program to Stimulate Competitive Research DE-SC0024161; U.S. Department of Energy’s Earth and Environmental System Modeling DE-SC0025208; and
610 NASA Precipitation Measurement Mission Science Team Grant 80NSSC22K0609. ILM was supported by the NOAA cooperative agreement NA22OAR4320151, for the Cooperative Institute for Earth System Research and Data Science (CIESRDS). The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.

References

- 615 Ackerman, A. S., Kirkpatrick, M. P., Stevens, D. E., and Toon, O. B.: The impact of humidity above stratiform clouds on indirect aerosol climate forcing, *Nature*, 432, 1014–1017, <https://doi.org/10.1038/nature03174>, 2004.
- Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-L., et al.: Bounding global aerosol radiative forcing of climate change, *Reviews of Geophysics*, 58, e2019RG000660, 2020.
- Bergeron, T.: On the physics of clouds and precipitation, *Proc. 5th Assembly UGGI, Lisbon, Portugal, 1935*, pp. 156–180, 1935.
- 620 Bosilovich, M., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A., Norris, P., et al.: Technical report series on global modeling and data assimilation, volume 43 merra-2: Initial evaluation of the climate, NASA-GMAO. Retrieved from <https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf>, 2015.
- Carslaw, K., Lee, L., Reddington, C., Pringle, K., Rap, A., Forster, P., Mann, G., Spracklen, D., Woodhouse, M., Regayre, L., et al.: Large contribution of natural aerosols to uncertainty in indirect forcing, *Nature*, 503, 67–71, 2013.
- 625 Carslaw, K. S., Gordon, H., Hamilton, D. S., Johnson, J. S., Regayre, L. A., Yoshioka, M., and Pringle, K. J.: Aerosols in the pre-industrial atmosphere, *Current Climate Change Reports*, 3, 1–15, 2017.
- Clarke, A. D., Owens, S. R., and Zhou, J.: An ultrafine sea-salt flux from breaking waves: Implications for cloud condensation nuclei in the remote marine atmosphere, *Journal of Geophysical Research: Atmospheres*, 111, 2006.
- Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., Emmons, L., Fasullo, J., Garcia, R., Gettelman, A., et al.: The community earth system model version 2 (CESM2), *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001916, 2020.
- 630 Eidhammer, T., Gettelman, A., Thayer-Calder, K., Watson-Parris, D., Elsaesser, G., Morrison, H., van Lier-Walqui, M., Song, C., and McCoy, D.: An extensible perturbed parameter ensemble (PPE) for the community atmosphere model version 6, *EGUsphere*, 2024, 1–27, 2024.
- Field, P. and Heymsfield, A.: Importance of snow to global precipitation, *Geophysical Research Letters*, 42, 9512–9520, 2015.
- 635 Field, P. R. and Furtado, K.: How biased is aircraft cloud sampling?, *Journal of Atmospheric and Oceanic Technology*, 33, 185–189, 2016.
- Forster, P. M.: Inference of climate sensitivity from analysis of Earth’s energy budget, *Annual Review of Earth and Planetary Sciences*, 44, 85–106, 2016.
- Fossum, K. N., Ovadnevaite, J., Ceburnis, D., Dall’Osto, M., Marullo, S., Bellacicco, M., Simó, R., Liu, D., Flynn, M., Zuend, A., et al.: Summertime primary and secondary contributions to Southern Ocean cloud condensation nuclei, *Scientific reports*, 8, 13844, 2018.
- 640 Gettelman, A. and Morrison, H.: Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes, *Journal of Climate*, 28, 1268–1287, 2015.
- Gettelman, A., Morrison, H., Santos, S., Bogenschutz, P., and Caldwell, P.: Advanced two-moment bulk microphysics for global models. Part II: Global model solutions and aerosol–cloud interactions, *Journal of Climate*, 28, 1288–1307, 2015.
- Gettelman, A., Bardeen, C., McCluskey, C. S., Järvinen, E., Stith, J., Bretherton, C., McFarquhar, G., Twohy, C., D’Alessandro, J., and Wu, W.: Simulating observations of Southern Ocean clouds and implications for climate, *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032619, 2020.
- 645 Ghan, S., Wang, M., Zhang, S., Ferrachat, S., Gettelman, A., Griesfeller, J., Kipling, Z., Lohmann, U., Morrison, H., Neubauer, D., et al.: Challenges in constraining anthropogenic aerosol effects on cloud radiative forcing using present-day spatiotemporal variability, *Proceedings of the National Academy of Sciences*, 113, 5804–5811, 2016.

- 650 Grosvenor, D. and Wood, R.: The effect of solar zenith angle on MODIS cloud optical and microphysical retrievals within marine liquid water clouds, *Atmospheric Chemistry and Physics*, 14, 7291–7321, 2014.
- Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R., Cairns, B., Chiu, J. C., Christensen, M., et al.: Remote sensing of droplet number concentration in warm clouds: A review of the current state of knowledge and perspectives, *Reviews of Geophysics*, 56, 409–453, <https://doi.org/https://doi.org/10.1029/2017RG000593>, 2018.
- 655 Gryspeerdt, E., Quaas, J., and Bellouin, N.: Constraining the aerosol influence on cloud fraction, *Journal of Geophysical Research: Atmospheres*, 121, 3566–3583, 2016.
- Gryspeerdt, E., McCoy, D. T., Crosbie, E., Moore, R. H., Nott, G. J., Painemal, D., Small-Griswold, J., Sorooshian, A., and Ziemba, L.: The impact of sampling strategy on the cloud droplet number concentration estimated from satellite data, *Atmospheric Measurement Techniques*, 15, 3875–3892, 2022.
- 660 Hamilton, D. S., Lee, L. A., Pringle, K. J., Reddington, C. L., Spracklen, D. V., and Carslaw, K. S.: Occurrence of pristine aerosol environments on a polluted planet, *Proceedings of the National Academy of Sciences*, 111, 18 466–18 471, 2014.
- Hudson, J. G. and Svensson, G.: Cloud microphysical relationships in California marine stratus, *Journal of Applied Meteorology* (1988–2005), pp. 2655–2666, 1995.
- Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Lee, L. A., Sexton, D. M., Rostron, J. W., Booth, B. B., and Carslaw, K. S.: The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing, *Atmospheric Chemistry and Physics*, 18, 13 031–13 053, 2018.
- 665 Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Turnock, S. T., Browse, J., Sexton, D. M., Rostron, J. W., Schutgens, N. A., Partridge, D. G., et al.: Robust observational constraint of uncertain aerosol processes and emissions in a climate model and the effect on aerosol radiative forcing, *Atmospheric Chemistry and Physics*, 20, 9491–9524, 2020.
- 670 Kang, L., Marchand, R., Wood, R., and McCoy, I. L.: Coalescence scavenging drives droplet number concentration in southern ocean low clouds, *Geophysical Research Letters*, 49, e2022GL097 819, 2022.
- Khairoutdinov, M. and Kogan, Y.: A new cloud physics parameterization in a large-eddy simulation model of marine stratocumulus, *Monthly weather review*, 128, 229–243, 2000.
- King, M. D., Platnick, S., Menzel, W. P., Ackerman, S. A., and Hubanks, P. A.: Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites, *IEEE transactions on geoscience and remote sensing*, 51, 3826–3852, 2013.
- 675 Klein, S. A. and Hall, A.: Emergent constraints for cloud feedbacks, *Current climate change reports*, 1, 276–287, 2015.
- Lance, S., Brock, C., Rogers, D., and Gordon, J.: Water droplet calibration of the Cloud Droplet Probe (CDP) and in-flight performance in liquid, ice and mixed-phase clouds during ARCPAC, *Atmospheric Measurement Techniques*, 3, 1683–1706, 2010.
- Lee, L., Carslaw, K., Pringle, K., Mann, G., and Spracklen, D.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, *Atmospheric Chemistry and Physics*, 11, 12 253–12 273, 2011.
- 680 Mallet, M. D., Miljevic, B., Humphries, R. S., Mace, G. G., Alexander, S. P., Protat, A., Chambers, S., Cravigan, L., DeMott, P. J., Fiddes, S., et al.: Biological enhancement of cloud droplet concentrations observed off East Antarctica, *npj Climate and Atmospheric Science*, 8, 113, 2025.
- McCluskey, C. S., Gettelman, A., Bardeen, C. G., DeMott, P. J., Moore, K. A., Kreidenweis, S. M., Hill, T. C., Barry, K. R., Twohy, C. H., Toohey, D. W., et al.: Simulating Southern Ocean aerosol and ice nucleating particles in the Community Earth System Model version 2, *Journal of Geophysical Research: Atmospheres*, 128, e2022JD036 955, 2023.

- McCoy, D. T., Burrows, S. M., Wood, R., Grosvenor, D. P., Elliott, S. M., Ma, P.-L., Rasch, P. J., and Hartmann, D. L.: Natural aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo, *Science advances*, 1, e1500 157, 2015.
- 690 McCoy, D. T., Bender, F. A.-M., Grosvenor, D. P., Mohrmann, J. K., Hartmann, D. L., Wood, R., and Field, P. R.: Predicting decadal trends in cloud droplet number concentration using reanalysis and satellite data, *Atmospheric Chemistry and Physics*, 18, 2035–2047, 2018.
- McCoy, I. L., McCoy, D. T., Wood, R., Regayre, L., Watson-Parris, D., Grosvenor, D. P., Mulcahy, J. P., Hu, Y., Bender, F. A.-M., Field, P. R., et al.: The hemispheric contrast in cloud microphysical properties constrains aerosol forcing, *Proceedings of the National Academy of Sciences*, 117, 18 998–19 006, 2020.
- 695 McCoy, I. L., Bretherton, C. S., Wood, R., Twohy, C. H., Gettelman, A., Bardeen, C. G., and Toohey, D. W.: Influences of recent particle formation on Southern Ocean aerosol variability and low cloud properties, *Journal of Geophysical Research: Atmospheres*, 126, e2020JD033 529, 2021.
- McFarquhar, G. M., Bretherton, C. S., Marchand, R., Protat, A., DeMott, P. J., Alexander, S. P., Roberts, G. C., Twohy, C. H., Toohey, D., Siems, S., et al.: Observations of clouds, aerosols, precipitation, and surface radiation over the Southern Ocean: An overview of CAPRICORN, MARCUS, MICRE, and SOCRATES, *Bulletin of the American Meteorological Society*, 102, E894–E928, 2021.
- 700 Mohrmann, J., Wood, R., McGibbon, J., Eastman, R., and Luke, E.: Drivers of seasonal variability in marine boundary layer aerosol number concentration investigated using a steady state approach, *Journal of Geophysical Research: Atmospheres*, 123, 1097–1112, 2018.
- Regayre, L. A., Johnson, J. S., Yoshioka, M., Pringle, K. J., Sexton, D. M., Booth, 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, *Atmospheric Chemistry and Physics*, 18, 9975–10 006, 2018.
- 705 Regayre, L. A., Schmale, J., Johnson, J. S., Tatzelt, C., Baccarini, A., Henning, S., Yoshioka, M., Stratmann, F., Gysel-Beer, M., Grosvenor, D. P., et al.: The value of remote marine aerosol measurements for constraining radiative forcing uncertainty, *Atmospheric Chemistry and Physics*, 20, 10 063–10 072, 2020.
- Regayre, L. A., Deaconu, L., Grosvenor, D. P., Sexton, D. M., Symonds, C., Langton, T., Watson-Paris, D., Mulcahy, J. P., Pringle, K. J., Richardson, M., et al.: Identifying climate model structural inconsistencies allows for tight constraint of aerosol radiative forcing, *Atmospheric Chemistry and Physics*, 23, 8749–8768, 2023.
- 710 Schutgens, N., Dubovik, O., Hasekamp, O., Torres, O., Jethva, H., Leonard, P. J., Litvinov, P., Redemann, J., Shinozuka, Y., De Leeuw, G., et al.: AEROCOM/AEROSAT AAOT & SSA study, part I: evaluation and intercomparison of satellite measurements, *Atmospheric Chemistry and Physics Discussions*, 2020, 1–35, 2020.
- Seinfeld, J. H. and Pandis, S. N.: *Atmospheric chemistry and physics: from air pollution to climate change*, John Wiley & Sons, 2016.
- 715 Silber, I., Jackson, R. C., Fridlind, A. M., Ackerman, A. S., Collis, S., Verlinde, J., and Ding, J.: The Earth Model Column Collaboratory (EMC 2) v1. 1: an open-source ground-based lidar and radar instrument simulator and subcolumn generator for large-scale models, *Geoscientific Model Development*, 15, 901–927, 2022.
- Song, C., McCoy, D. T., Eidhammer, T., Gettelman, A., McCoy, I. L., Watson-Parris, D., Wall, C. J., Elsaesser, G., and Wood, R.: Buffering of aerosol-cloud adjustments by coupling between radiative susceptibility and precipitation efficiency, *Geophysical Research Letters*, 51, e2024GL108 663, 2024.
- 720 Twomey, S.: The influence of pollution on the shortwave albedo of clouds, *Journal of the atmospheric sciences*, 34, 1149–1152, 1977.
- Wall, C. J., Norris, J. R., Possner, A., McCoy, D. T., McCoy, I. L., and Lutsko, N. J.: Assessing effective radiative forcing from aerosol–cloud interactions over the global ocean, *Proceedings of the National Academy of Sciences*, 119, e2210481 119, 2022.

Watson-Parris, D. and Smith, C. J.: Large uncertainty in future warming due to aerosol forcing, *Nature Climate Change*, 12, 1111–1113, 2022.

Watson-Parris, D., Williams, A., Deaconu, L., and Stier, P.: Model calibration using ESEm v1. 1.0—an open, scalable Earth system emulator, *Geoscientific Model Development*, 14, 7659–7672, 2021.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K.: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble, *Climate dynamics*, 41, 1703–1729, 2013.

Wood, R.: Rate of loss of cloud droplets by coalescence in warm clouds, *Journal of Geophysical Research: Atmospheres*, 111, 2006.

Wood, R.: Stratocumulus clouds, *Monthly Weather Review*, 140, 2373–2423, 2012.

Wood, R., Leon, D., Lebsock, M., Snider, J., and Clarke, A. D.: Precipitation driving of droplet concentration variability in marine low clouds, *Journal of Geophysical Research: Atmospheres*, 117, 2012.

Zhang, Z., Werner, F., Cho, H.-M., Wind, G., Platnick, S., Ackerman, A., Di Girolamo, L., Marshak, A., and Meyer, K.: A framework based on 2-D Taylor expansion for quantifying the impacts of subpixel reflectance variance and covariance on cloud optical thickness and effective radius retrievals based on the bispectral method, *Journal of Geophysical Research: Atmospheres*, 121, 7007–7025, 2016.

Zheng, G., Wang, Y., Aiken, A. C., Gallo, F., Jensen, M. P., Kollias, P., Kuang, C., Luke, E., Springston, S., Uin, J., et al.: Marine boundary layer aerosol in the eastern North Atlantic: seasonal variations and key controlling processes, *Atmospheric Chemistry and Physics*, 18, 17 615–17 635, 2018.

Zheng, X., Dong, X., Xi, B., Logan, T., and Wang, Y.: Distinctive aerosol–cloud–precipitation interactions in marine boundary layer clouds from the ACE-ENA and SOCRATES aircraft field campaigns, *Atmospheric Chemistry and Physics*, 24, 10 323–10 347, 2024.

Zhou, X., Atlas, R., McCoy, I. L., Bretherton, C. S., Bardeen, C., Gettelman, A., Lin, P., and Ming, Y.: Evaluation of cloud and precipitation simulations in CAM6 and AM4 using observations over the Southern Ocean, *Earth and Space Science*, 8, e2020EA001 241, 2021.