We thank both reviewers and the editor for their time and thought in reviewing our paper. Reviewer comments are in black and our responses are in red. We are using the latexdiff to track changes and provide screenshots of changed text for the reviewers' convenience as well as a track changes version of the manuscript. The line numbers in the response refer to the track changes file.

Reviewer #1

The authors use field measurements from the SOCRATES campaign to constrain a CAM6 perturbed parameter ensemble. With a primary focus on CCN and Nd, the authors use an emulator to create surrogate models and constrain plausible parameter combination. Lastly, the authors show how observational uncertainty affects the ability to constrain.

The paper is well written and the figures provide sufficient visual context. There are a few concerns that the authors should address before publication.

Major concerns

The authors stress in several places that the performance of the emulator is crucial for the task at hand. Looking at Fig. 7, the colors of points and shading strongly disagree in many places. I wonder if the authors have an explanation of why the apparent performance is so poor and whether a better emulator is needed (or even possible).

Response: Thank you for pointing this out. We agree that the colors of the points do not always align with the color shading in Figure 7. This is because the color shading represents 2D bin averages of global mean present-day (PD) Nd (Figure 7a) and global mean ΔNd (Figure 7b), while the color of the points corresponds to individual PPE members without any averaging. Bin-averaged shading tends to smooth out extreme values, whereas individual points can reflect larger variability. Therefore, it is not surprising to observe some mismatch between the two.

However, the overall color patterns (color gradiant) between the shading and the points show good agreement. Moreover, the emulator performance for both PD Nd and ΔNd, as shown in Figure S2, is strong: the **emulator's mean predictions** closely follow the 1-to-1 line when compared with testing data. This gives us confidence in the emulator's ability to accurately represent the underlying model behavior, and we do not believe a different emulator is necessary in this case, as Figure 7 uses **the emulator mean prediction**.

Emulator uncertainty becomes important primarily when applying observational constraints to the PPE. We have acknowledged this in the revised manuscript, and it is examined in detail in Figures 8b and 9.

To clarify the mismatch between the color points and the color shading in Figure 7, we have added more explanation of the color shading and point coloring in the caption of Figure 7 in the revised manuscript.

The color scale in Figure 7 has been revised from the rainbow scheme to a palette that is accessible to colorblind readers.

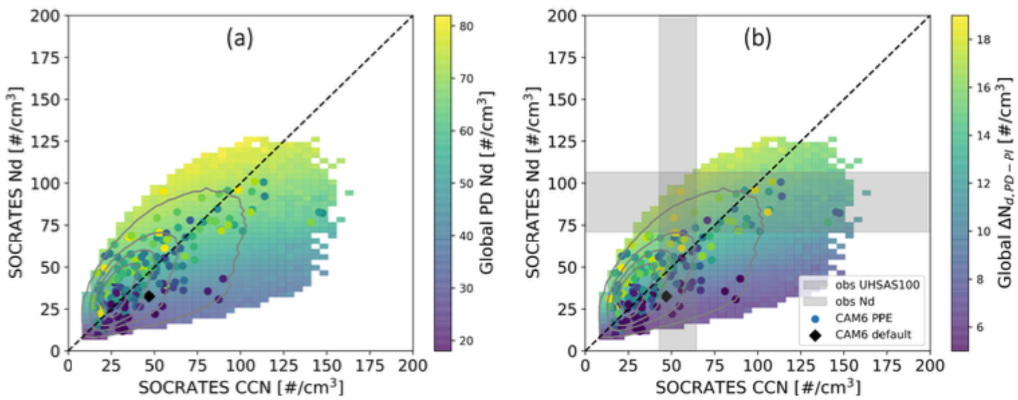


Figure 7. (a) SOCRATES campaign-mean N_d versus campaign-mean CCN and colored by present-day N_d from the CAM6 PPE members (color dots) and 1M emulations from the PPE (color shading). Emulate density is shown in solid contours. (~~d~~b) The same with (~~e~~a) but colored by $\Delta N_{d,PD-PI}$. The 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 Nd 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 ±20% from the campaign-mean.

The authors largely leave out cloud macrophysical properties (e.g., cloud fraction, cloud geometric thickness, etc.). Does it go without saying that the nudged PPE runs produce plausible macrophysical properties? The authors should at least provide a brief assessment.

Response: Thank you for the suggestion. The main focus of this study is to investigate constraints on cloud microphysical properties using aircraft in-situ measurements. While a detailed evaluation of cloud macrophysical properties is beyond the scope of this work, we agree that it is important to establish the credibility of the nudged PPE simulations. Previous studies using similar CAM6 nudged configurations have shown that these simulations produce a wide spread in present-day cloud microphysical (Nd) and macrophysical (LWP) properties, with mean-state values falling within the observational range derived from satellite remote sensing (Song et al., 2024). In addition, CAM6 simulations along aircraft flight tracks using the default parameter configuration reproduce key features of in-situ observations, including cloud phase, cloud location, and boundary layer structure (Gettelman et al., 2020). We now briefly note this in Line 114-120 in the revised manuscript and include citations to support it.

Previous studies have shown that the CAM6 PPE, configured with 2-year global simulations, produces a wide spread in 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.

Along with the above concern, I am wondering if this study is limited to stratiform clouds (as no convective scheme is described in Sec. 2). Have the authors ensured that all clouds in CAM6 are stratiform over the SOCRATES domain? How would the results change if a substantial portion was handled by the convective scheme?

Response: Yes, this study is restricted to low-level, liquid clouds. We focus our analysis on aircraft measurements of aerosol and cloud properties below 2 km, which correspond to the marine boundary layer and exclude clouds formed by deep convection.

In CAM6, cloud microphysics, including cloud droplet number concentration (Nd), is handled exclusively by the stratiform (large-scale) scheme (MG2). The convective scheme does not calculate microphysical properties such as Nd. Therefore, all cloud variables used in this analysis are from stratiform clouds.

The authors use observed surface precipitation rates, but it is unclear where these rates stem from. The authors need to update Section 2 and describe the retrieval.

Response: Thank you for pointing this out. We apologize for the lack of clarification. In this study, we did not use **observed** surface precipitation rates. As noted in the manuscript, making a direct (apples-to-apples) comparison between CAM6 precipitation and radar–lidar-retrieved cloud-base precipitation is challenging due to mismatch in vertical level at which precipitation is reported, as well as retrieval uncertainty and instrument sensitivity. Instead, we conducted a sensitivity test using two **hypothetical** campaign-mean surface precipitation rates to examine how such constraints would influence ΔNd. We have now clarified this approach in the Results section of the manuscript (Line 486-488, 569-570), where it was discussed.

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 ~~would be if we know~~ 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

examine what the constraints would be like if we know the observed surface precipitation. However, the constraints from two hypothetical surface precipitation is minimal (Figure S8). We did not include cloud base precipitation rate as a constraint due

Minor concerns

l. 16 “possible” rather than “much easier”

Response. Corrected.

condense onto. CCN make cloud droplet formation ~~much easier~~ possible in atmospheric conditions. Aerosols from anthro-

ll. 105ff Please briefly describe the synoptic situation encountered during the flights.

Response: Thank you for the suggestion. We have now added a brief description of the synoptic situation in the revised manuscript (Line 106-108). The campaign took place in the midlatitude Southern Ocean during austral summer and was characterized by frequent passages of frontal systems, postfrontal stratocumulus decks, and cyclonic activity typical of the storm track region (McFarquhar et al., 2021).

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-

ll. 265-266 Could a lower updraft speed also explain this?

Response: This is a great point—thank you for raising it. Our analysis supports the interpretation that subgrid-scale vertical velocity is underestimated in CAM6, which could help explain the low-biased Nd. We have clarified this point in the revised Results section. (Line 286, 350-352) 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.

transport and in generating supersaturation~~, which are important for aerosol activation~~.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.

Typo(s)

l. 453 “of without”

response: we agree the original sentence is confusing. Corrected.

examined their joint effects on $\Delta N_{d,PD-PI}$ in Figure 8. We found great improvement on the N_d constraint when including the effects of CCN~~under condition of without~~ , assuming no emulator uncertainty (Figure 8b). The $\Delta N_{d,PD-PI}$ range is narrowed

Reviewer #2

This is a review of ‘Aircraft In-situ Measurements from SOCRATES Constrain Anthropogenic Perturbations of Cloud Droplet Number’ by Song et al. It is an important study that warrants publication in ACP. The analysis is very in-depth and complex. Using Southern Ocean observations to constrain global cloud properties using a perturbed parameter ensemble is never going to be an easy thing to synthesize, and I did have to reread various sections to absorb everything. But I have learned a lot in the process and I think overall this is a great study.

I have a major comment, related to my public comment about the use of N100 as a proxy for CCN, as well as a number of minor and technical comments. I suggest major revisions because I think it is very important that the conclusion that “CCN” observations do not provide any constraint on global Nd is revisited. In reality, I do not think the major revision should take long to address.

Major comments:

The authors use observations of N100 (particle concentration above 100nm) as a proxy for CCN. They cite another study from the same observational aircraft dataset showing a slope of ~1:1 between CCN (at 0.2% supersaturation) and N100. The authors responded to my public comment regarding this, stating that it is the distribution mean that is used as the observational constraint. I understand the logic of this, but it should be noted that the ratio of campaign mean CCN0.2:N100 in McCoy et al. (2021) Figure S2 is ~1.08 (+/- ~0.3). If we take that average, that means the use of N100 is underestimating CCN0.2 by ~8%. This is a potential known systematic error that is not discussed (e.g. Lines 165 to 175). What impact would this have on the analysis for the observational constraint on the distribution of PPE emulated global PI - PD Nd? I think this could be roughly determined without needing to repeat the analysis with the actual CCN0.2 observational dataset. I am trying to assess the potential impact of this by looking at Figure 7b if the vertical grey shaded area (representing observed N100) was shifted to the right. It's difficult without the raw data, but I think it would shift the distribution of plausible emulations constrained with only CCN observations of PI - PD Nd to higher values. It's difficult for me to glean from the density contours and colour scale whether the distribution would be narrower or not. It's clear that the Nd-only constraint will still be much stronger, but it might also slightly change the Nd + CCN constraint.

The activation diameter for Southern Ocean aerosol is likely always going to be less than 100nm for supersaturations of 0.2%, and likely around 80nm for the aerosol population sampled during SOCRATES (e.g. see Fossum et al., 2018; Figure 2b, where mP air masses contain aerosol with similar characteristics to oceanic air masses south of Australia (Mallet et al., 2025, Figure 2b)). I suspect that the SOCRATES CCN0.2:N100 ratio of ~1.08 shown in McCoy et al 2021 Figure S2a is likely biased low due to some data points where this ratio is close to zero, which is physically implausible. I therefore think an 8% underestimation of CCN is conservative, and it would be interesting to see what the impact would be on Figure 7 and 8 and associated analyses if a ratio closer to the upper end of the error bar (~1.4) in McCoy et al., 2021 Figure S2 was used.

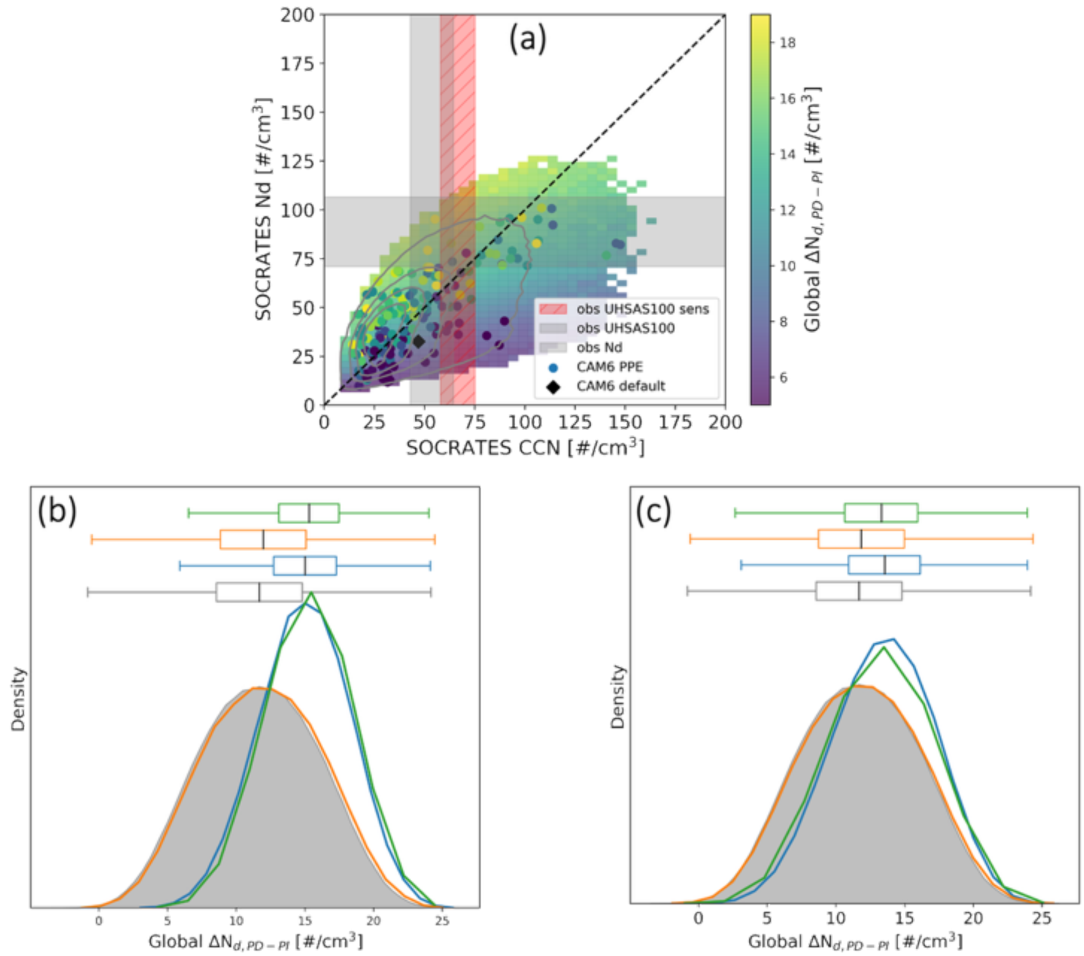
I recognise the huge amount of effort that's already gone into this paper. I also recognise that it is likely that the N100 data is probably more readily available and processed for comparison to the model output. Ideally the whole analysis would be repeated for the actual QA'd CCN0.2, but that might not be feasible. At the very least I think the authors should test the impact that increasing the campaign mean observed N100 by 8% and 40% so that it aligns with the campaign mean observed CCN0.2. I think that an increase in the "observed" CCN by 8-40% would probably change the population of plausible PPEs and emulations, which would change the constrained parameter spaces and constrained PD - PI Nd distributions. This would impact Figure 3a/c, Figure 7b, Figure 8, and Figure 9, as well as many of the conclusion and discussion statements regarding the CCN constraint. If those two tests change nothing, then only some of the discussion might need expanding so that other readers with similar thoughts to me understand.

In response to my public comment about the use of N100 as a proxy for CCN0.2, the authors stated "we use N100 as a proxy for CCN to enable comparison across multiple CAM6 simulations with varied parameter combinations. One advantage of using N100 is that it allows direct comparison of our perturbed parameter ensemble (PPE) results with those from previous studies (e.g., Figure 2)." My understanding is that the CAM6 model used for the PPE work uses information from simulated aerosol to calculate an activation diameter for a particular supersaturation (I think from my reading the Abdul-Razzak & Ghan activation scheme with κ-Köhler is used). So it isn't entirely clear why using observations of N100 is a direct advantage for model comparison over using actual CCN observations. Observations of N100 would be the better choice if CAM6 used a static 100nm activation diameter, but then the use of the term CCN would be misleading in this study and the conclusion would be that observations of N100 alone do not constrain estimates of global PI Nd.

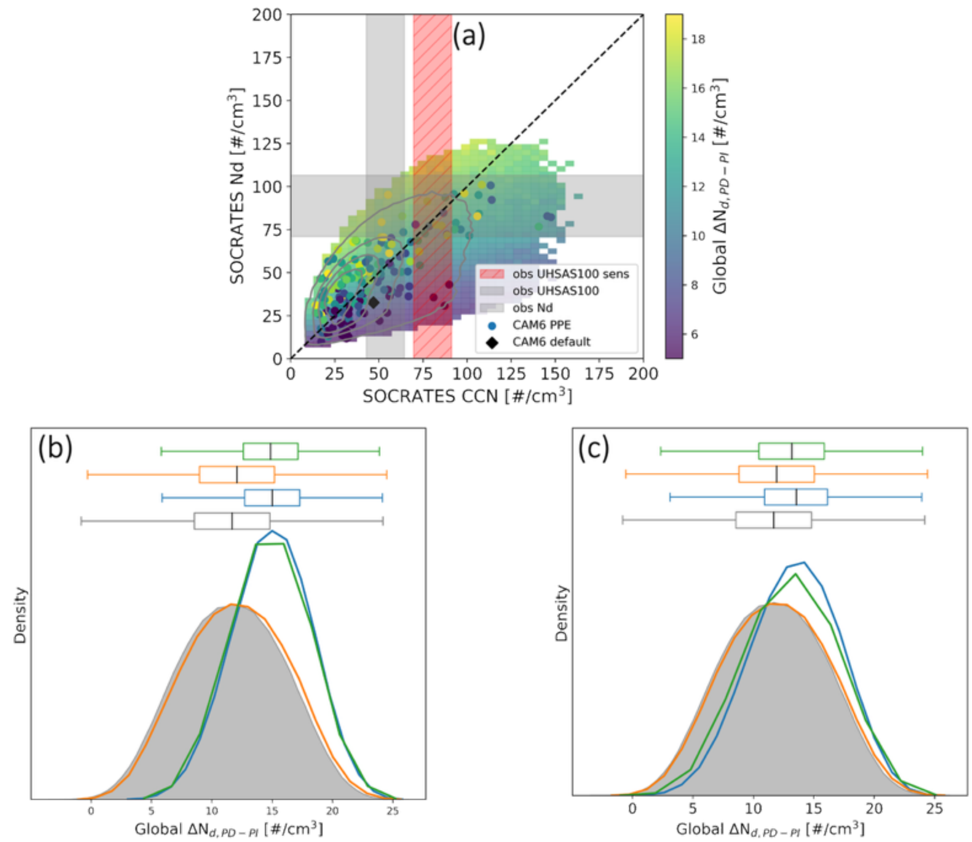
This might seem nitpicking, but it does seem to me that a slight underestimation of CCN by using N100 as a proxy could change part of the outcome of the paper. The value of Nd observations over the Southern Ocean for global climate modelling has been thoroughly demonstrated here. If CCN observations from the pristine Southern Ocean only provide a constraint on global estimates of the impact of anthropogenic aerosols on cloud microphysics globally when collected in conjunction with Nd observations (which are arguably more expensive and logistically challenge to collect), then that might (and should) influence decision making and justification for future measurement efforts in the Southern Ocean. This paper is potentially quite impactful, so I do think it is worth revisiting the potential implications of using N100 as a proxy for CCN while considering even small differences in the unity of the relationship between these two.

Response: Thank you for the detailed and thoughtful comment. We agree that using UHSAS100 campaign-mean values as a proxy for CCN at 0.2% supersaturation (CCN0.2) may lead to a systematic underestimation of CCN, as noted in (McCoy et al., 2021). In response, we conducted sensitivity tests to assess how this potential bias could impact our constraints on the PD–PI change in Nd. We provide a brief discussion in the response file, while the detailed discussion can be found in the main text (Line 186).

To reflect the potential offset between UHSAS100 and CCN0.2 *we increased the observed "CCN" by 8% and 40%, representing the lower and upper bounds of the* CCN0.2: N100 ratio uncertainty. This sensitivity scenario is now illustrated in the revised Figure S9, where the additional red shaded bars represent the expanded observational constraint range. The result suggests the constraints on PD-PI change in Nd does not shift much as well in this case.



In the second sensitivity test, we further increase the observed ‘CCN’ to be 0.3-0.7 higher than the campaign-mean UHSAS100. The resulting constraints on the PD–PI change in Nd also do not shift significantly in this case.



Although an increased 'CCN' does not change our results, we agree that this analysis should highlight the potential issue of using UHSAS100 as a proxy for CCN0.2. We have added further discussion on this point and included Figure S9 as a sensitivity experiment (Line186-194, 496-499).

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 ($\pm$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 USHS100 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).

cloud base precipitation rate to 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).

Minor comments:

The use of CCN and Nd separately and combined to constrain the PPEs is interesting. Could the authors add a few sentences to the conclusion about future directions? Are there any other observable quantities that could also be used given enough measurements (e.g. aerosol composition, aerosol size, ice nucleating particles, cloud ice crystal concentration)? I could imagine a scenario whereby introducing more observational constraints severely limits or even eliminates the number of plausible PPEs/emulations, but I wonder if that in itself could be used to highlight structural issues within these models.

Response: Thank you for the thoughtful comment. As noted in Section 3.4, the improvement in constraints from using CCN and Nd together occurs primarily when observational uncertainty is small. This highlights the importance of obtaining accurate airborne measurements of both variables. In future campaigns, we suggest optimizing sampling strategies (e.g., stochastic sampling) to enable more direct comparisons with large-scale Earth system models. This is already discussed in Line 584-587.

In terms of future directions, we agree that additional observable quantities could further constrain the PPE and help identify structural model limitations. For example, in-situ measurements of aerosol composition (e.g., sulfate or organic fraction) and aerosol size distributions could be incorporated. CAM6 provides mass and number concentrations by mode and species, allowing for some level of comparison, though differences in mixing assumptions and humidification must be considered. These added constraints could further narrow the space of plausible PPE configurations, or, if they provide weaker constraint, they may instead expose structural model uncertainties.

Incorporating cloud-phase-relevant quantities, 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, we agree that they represent promising directions for future work.

Finally, while our study uses only in-situ measurements, future studies could integrate lidar–radar-retrieved variables to enhance constraints. We also suggest further development of instrument simulators that process model output under observation-like conditions, allowing for more direct comparisons.

We have added a discussion of these ideas in the revised manuscript (Line 588-596).

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, 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, they represent

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

There appear to be some large differences between the PPE members and the emulations. The other review raised this concern. I do not have the capacity to thoroughly review this aspect of the analyses, but I do think the authors have acknowledged potential limitations with the emulations and gone to lengths to consider the impact of the emulator uncertainty (e.g. Figure 8b, Figure 9).

Response: Thanks for bring this up. The apparent discrepancies likely stem from Figure 7. The mismatch between the PPE members and the emulator output arises from how the data are plotted. As noted in our response to Reviewer 1, the color shading represents 2D bin-averaged values of global mean present-day (PD) Nd (Figure 7a) and global mean ΔNd (Figure 7b), while the colored points represent values from individual PPE members without any averaging. Bin-averaging smooths out extreme values, whereas individual points can reflect larger variability. Therefore, some degree of mismatch between the two is expected. We have brought this point in the revised manuscript.

Figure 7b shows the emulator mean prediction. The emulator performs well for both PD Nd and ΔNd, as demonstrated in Figure S2, and this is reflected in the good agreement in color gradiant between the color shading and the data points in Figure 7b. This gives us confidence in the emulator’s ability to capture the underlying model behavior, and we do not find it necessary to use a different emulator. Emulator uncertainty becomes important primarily when applying observational constraints, which we discuss in Figures 8b and 9.

We have clarified the differences in plotting of shaded areas and scatter points in the caption of Figure 7.

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. (~~d~~b) The same with (~~e~~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 Nd 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 ±20% from the campaign-mean.

Section 3.3.2 is quite complex. My understanding is that because there is a good relationship between the global Nd for the PI and PD (fig 4), it's important to demonstrate there is a physically plausible reason for that in order to strengthen the following results/conclusions about the use of present-day pristine observations to constrain PI global Nd. I haven't rigorously gone through the maths for this section due to time constraints, but the logic seems sound to me.

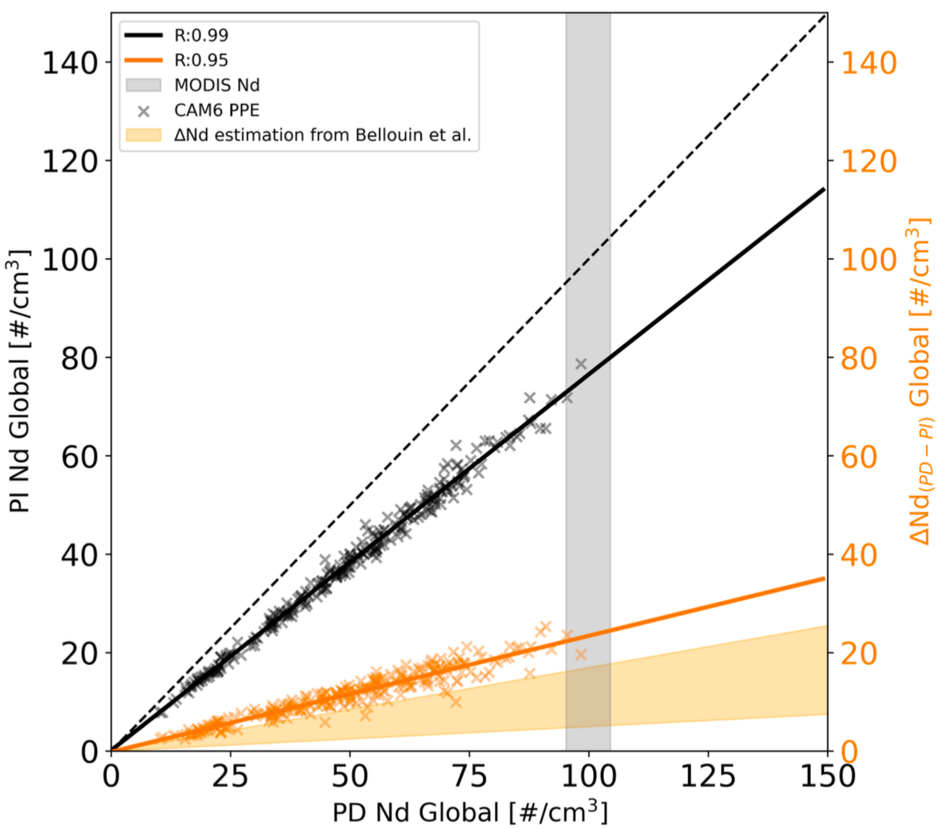
Response: You are correct. Section 3.3.2 uses math equations based on our understanding of sources and sinks of Nd/aerosol to explain the emergent relationships we find in Figure 4. Following that we observed CCN and Nd to constrain the ΔNd ranges from the PPE.

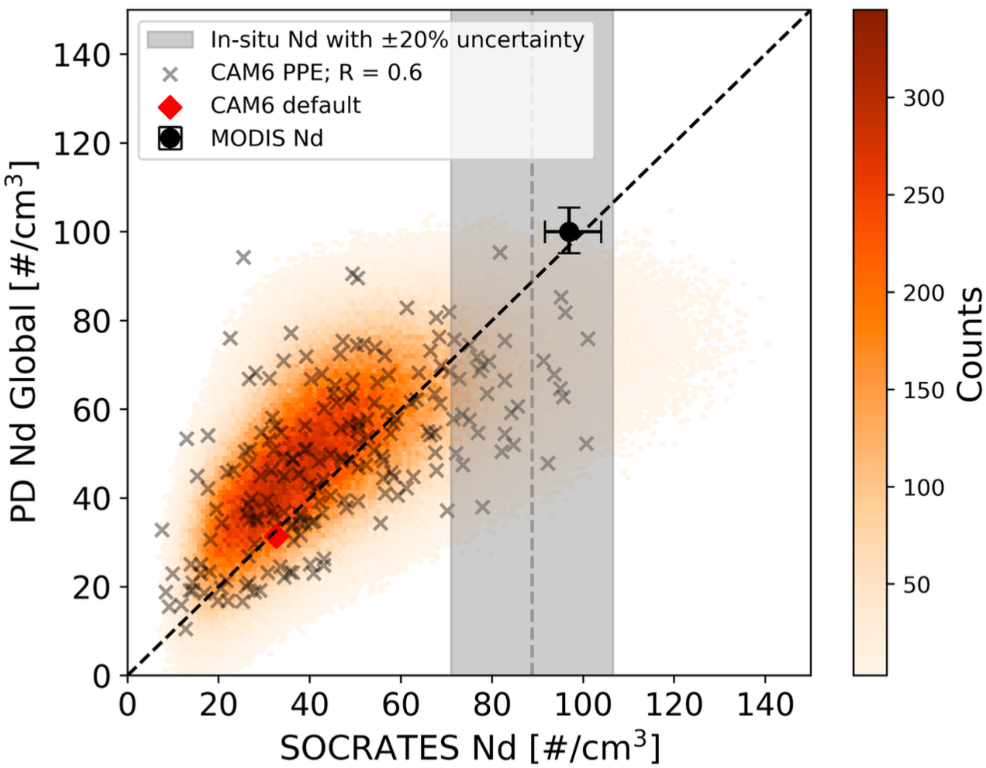
L247: Are these cases focused only on warm liquid clouds? I would have thought there was a significant occurrence of supercooled liquid in these flights.

Response: Thank you for the insightful comment. While our study focuses on liquid-phase clouds below 2 km, we acknowledge that supercooled liquid is common in the Southern Ocean during SOCRATES. In the manuscript (Section 3.2: line 335-344), we note that the SO is dominated by supercooled liquid clouds (Gettelman et al., 2020; McCluskey et al., 2023), and that glaciation via the Bergeron–Findeisen process may act as a sink for cloud droplet number concentration (Nd). However, we also explain that this effect is likely minimal in our study, as we restrict analysis to cloud layers below 2 km, where the majority of snow melts and contributes to rain [see Figure 2 in (Field & Heymsfield, 2015)]. The limited influence of ice-phase processes is also reflected in our process-based constraints shown in Figure S6, which includes the 2 km altitude restriction. This supports the assumption that the ice-phase sink of Nd is small within our filtered domain.

Figure 4 and 5: The grey shaded area in Figure 4 is stated to represent the 95% confidence on the interannual range of global oceanic mean Nd from MODIS (visually, ~80 - 115 cm^-3). The black dot in Figure 5 also shows this, but the error bar along the y-axis expressed a much tighter range, despite the caption stating it represents the same thing as Figure 4.

Response: Thank you for pointing this out. You are correct that there is an inconsistency between the shaded region in Figure 4 and the error bar in Figure 5. The wider range in Figure 4 is due to a mistake in how the uncertainty was calculated: we originally used the 2.5th and 97.5th percentiles of **monthly** global oceanic mean Nd values, rather than computing the interannual spread from **yearly means**. In contrast, the error bar in Figure 5 correctly reflects the 95% confidence interval calculated across **annual means**. We have corrected this in the revised Figure 4 and updated the caption to clarify the definition of the uncertainty range.





Technical comments:

I noticed a few typos throughout and have highlighted them where I could, but it would be worth a final proofread (including supplementary material) before the next submission.

Figure 1: I very much like this figure. Could the caption indicate whether this is annual, or austral summer?

Response: Thank you for bringing this up. Figure 1a shows simulations run over two years, while Figure 2a shows along-flight-track simulations run for January–March 2018, covering late austral summer into early autumn. We have clarified this point in the caption of Figure 1.

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.

Lines 354 - 358/Figure 1. Natural aerosols do indeed dominate the SO. But some of the differences between the PI and PD Nd could also be due to changes in non-aerosol drivers (e.g. precipitation). This is discussed later, but the logic could be brought forward earlier (in a brief sentence even), otherwise reading chronologically it might seem that this hasn't been considered.

Response: Thank you for the suggestion! We have now added a brief sentence in Line 379-381 to discussion the impact of non-aerosol drivers (i.e., precipitation scavenging) on Nd budget.

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.

Figure 2 were PPEs 010, 237, and 244 chosen at random as a demonstration of the different outputs? Or do they represent PPEs that resulted in good agreement (010, 237) and poor agreement (244) between the observed and simulated Nd?

Response: Thank you for the question. These PPE members were chosen to represent cases with both good and poor agreement between the simulated and observed **CCN and** Nd. For instance, PPE 010 and 244 (Figure 2b and 2d) show relatively good agreement in CCN but exhibit low-biased Nd. We have clarified this point in the revised caption of Figure 2 the main text.

to 1 μm from UHSAS100.

Figure 2 shows a subsample of ensemble members with varying levels of agreement with observations, but a positive cor-

Figure 2. Relationships between SOCRATES CCN and in-cloud cloud droplet number concentration (N_d) from in-situ measurements (red) and CAM6 members (black)~~from~~, based on flight composites along individual flight tracks (scatters). ~~Observations~~ Flight composite are ~~binned~~ 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 × 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 ≤ 1%, we set N_d = 0). ~~CAM6 PPE CCN and N_d are collocated to observations (50 m × 2 min bins) by linear interpolation for individual PPE members.~~ 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.

11

Figure 2: What each data point represents is not entirely clear. After reading further on and looking at Figure S1, I think the data points are indeed averages for individual flights, so I recommend changing "data are from individual flight tracks" to "data are averaged for each individual flight track" to make it clearer. Ideally some measure of spread/uncertainty would be expressed around those data points or stated where applicable.

Response: Thanks for pointing this out! We agree it was not clear in the original manuscript. The data points in Figure 2 represent flight composites along individual flight tracks, illustrating relationships between SOCRATES CCN and in-cloud cloud droplet number concentration (Nd) from in-situ measurements (red) and CAM6 members (black).

Flight composites are constructed by binning observations into 50 m (altitude) × 2 min (time) intervals for each flight. CAM6 PPE CCN and in-cloud Nd are collocated to the observational composites by linear interpolation for each PPE member. Bin medians are then taken for comparison with CAM6 models following (McCoy et al., 2021).

We have clarified this in the caption in Figure 2 in the revised manuscript.

Figure 2. Relationships between SOCRATES CCN and in-cloud cloud droplet number concentration (N_d) from in-situ measurements (red) and CAM6 members (black)~~from~~ , based on flight composites along individual flight tracks (scatters). ~~Observations~~ Flight composite are ~~binned~~ 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 × 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). ~~CAM6 PPE CCN and N_d are collocated to observations (50 m × 2 min bins) by linear interpolation for individual PPE members.~~ 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.

11

Line 288 “detail” should be “detailed”

Response: Corrected! (Line 307)

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 ~~detail~~ detailed parameter explanation is in Table S1). The sea salt

Line 292 “process” should be “processes”

Response: Corrected! (311)

2021; Zhou et al., 2021).

Constraints on precipitation ~~process~~ processes point to the importance of precipitation as an aerosol sink. One of the key

Line 378: redundant use of “can”.

Response: Deleted! (Line 401)

N_d sampled during SOCRATES contains information for globally-relevant processes (Figure 5), but do PD observations of aerosol and cloud properties ~~can~~ constrain the anthropogenic perturbation in N_d? We find this to be the case in the context of

Line 415. The last sentence in this paragraph doesn’t read properly. Maybe it should be a comma before “we”.

Response: Thanks for pointing this out! We agree the connection between the two sentences is not clear. We have corrected this in the revised manuscript (Line 441-445).

of CCN and N_d as discussed in secton 3.2. ~~We~~ Inspired by this, we wanted to examine the effects of ~~surface~~ 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 used, based on idealized sensitivity tests discussed in Section 3.3.3.

Figure 7 caption references (d) and (c) but there’s only (a) and (b). Furthermore, ideally the legend caption in the right panel would include the “,PD-PI”. The range for the grey shading indicating the observed Nd and UHSAS100 should be described in the figure caption. The rainbow colour scale isn’t colour blind or grayscale-printer friendly. I don’t know if editorial policy yet is to enforce colour-blind friendly colour scales, but if this figure ends up being redone, I’d encourage a different continuous colour scale.

Response: Thank you for pointing this out! We have corrected the figure caption and labels, and added descriptions for the grey shaded bars. All instances of “ΔNd” have been updated to “ΔNd,PD-PI” for consistency. We have also replaced the rainbow color scale with a color-blind–friendly alternative.

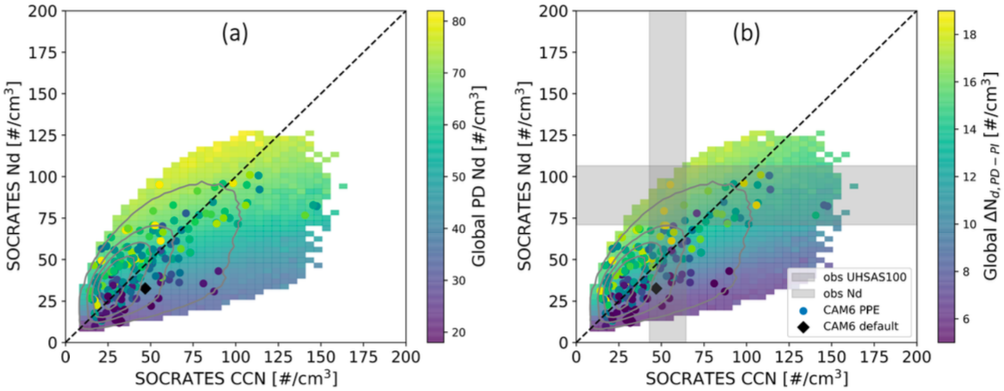


Figure 7. (a) SOCRATES campaign-mean N_d versus campaign-mean CCN and colored by present-day N_d from the CAM6 PPE members (color dots) and 1M emulations from the PPE (color shading). Emulate density is shown in solid contours. (~~d~~b) The same with (~~e~~a) but colored by $\Delta N_{d,PD-PI}$. The 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 Nd 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 ±20% from the campaign-mean.

Line 510: "(e.g. 262)" should be ("i.e. 262)"

Response: corrected! (Line 543)

While a large number of ensemble members (~~e.g.~~i.e ., 262) were integrated

Figure S5: Figure title has typo ("verus")

Response: corrected!

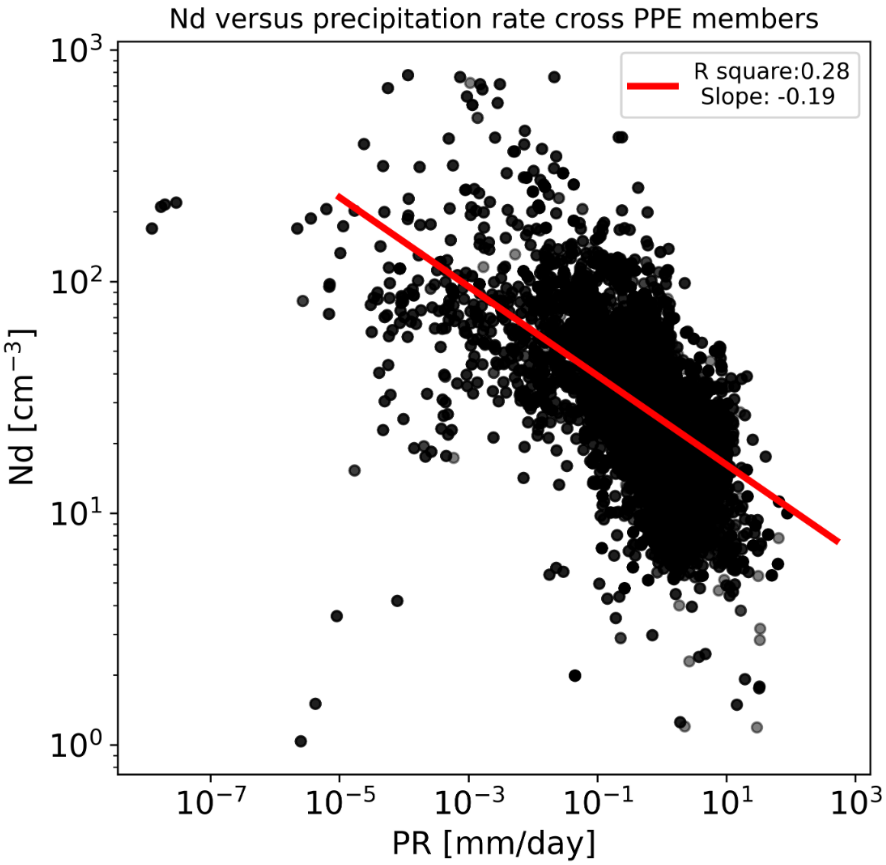


Figure S5: Nd versus precipitation rate. Each dot represents one simulation time step averaged across ensemble members.

References (for peer review discussion purposes only, no need to cite within text):

McCoy et al. (2021) – DOI: 10.1029/2020JD033529

Fossum et al. (2018) – DOI: 10.1038/s41598-018-32047-4

Mallet et al. (2025) – DOI: 10.1038/s41612-025-00990-5

Reference:

Field, P. R., & Heymsfield, A. J. (2015). Importance of snow to global precipitation. *Geophysical Research Letters*, *42*(21), 9512–9520. https://doi.org/10.1002/2015GL065497

Gettelman, A., Bardeen, C. G., McCluskey, C. S., Järvinen, E., Stith, J., Bretherton, C., et al. (2020). Simulating Observations of Southern Ocean Clouds and Implications for Climate. *Journal of Geophysical Research: Atmospheres*, *125*(21), e2020JD032619. https://doi.org/10.1029/2020JD032619

McCluskey, C. S., Gettelman, A., Bardeen, C. G., DeMott, P. J., Moore, K. A., Kreidenweis, S. M., et al. (2023). Simulating Southern Ocean Aerosol and Ice Nucleating Particles in the Community Earth System Model Version 2. *Journal of Geophysical Research: Atmospheres*, *128*(8), e2022JD036955. https://doi.org/10.1029/2022JD036955

McCoy, I. L., Bretherton, C. S., Wood, R., Twohy, C. H., Gettelman, A., Bardeen, C. G., & Toohey, D. W. (2021). Influences of Recent Particle Formation on Southern Ocean Aerosol Variability and Low Cloud Properties. *Journal of Geophysical Research: Atmospheres*, *126*(8), e2020JD033529. https://doi.org/10.1029/2020JD033529

McFarquhar, G. M., Bretherton, C. S., Marchand, R., Protat, A., DeMott, P. J., Alexander, S. P., et al. (2021). 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*(4), E894–E928. https://doi.org/10.1175/BAMS-D-20-0132.1

Song, C., McCoy, D. T., Eidhammer, T., Gettelman, A., McCoy, I. L., Watson-Parris, D., et al. (2024). Buffering of Aerosol-Cloud Adjustments by Coupling Between Radiative Susceptibility and Precipitation Efficiency. *Geophysical Research Letters*, *51*(11), e2024GL108663. https://doi.org/10.1029/2024GL108663