

Overall, the authors have put extensive work towards addressing my comments; and I thank the authors for doing so. However, some of my concerns still stand, particularly related to biases associated with aerosol measurements in-cloud. I recommend the authors address the concerns below. I respond to my previous three major concerns in the prior review below:

Addressing Major Concern #1:

The authors go through great lengths to argue large aerosol measurements (N_{aer500}) are not biased by ice crystals, which is much appreciated. But I am not convinced by their arguments. They note that most in-cloud samples have $N_{aer500}=0$. However, the frequencies of $N_{aer500}=0$ and >0 may correspond with low and high IWC, respectively. I had requested in Major Comment #2 that they should provide sensitivity tests for what I would argue to be more realistic in-cloud thresholds. These larger thresholds would most likely be associated with greater frequencies of $N_{aer500}>0$ due to the greater likelihood of shattering. They also mention small ice particles having comparable sizes to large aerosol have very low lifespans. This does not account for the possibility of ice shattering, where studies show relatively larger ice particles are more likely to shatter (references within Section 3 of McFarquhar et al. (2017)). Comments related to those associated with their respective supplementary figures are provided below:

Addressing figure R1: I initially recommended they provide multiple case studies showing N_{aer500} within and in the vicinity of the cloud. They provide one time series and state: “the fluctuations of both N_{a100} and N_{a500} are independent of those of IWC.” However, there is a clear increase in the frequency of $N_{aer500}>0$ where IWC peaks (~25:52:00). They then argue that N_{aer500} should be greatest where D_i is smallest and N_i is relatively high at 26:00:00, but no aerosols are observed at this time, thus supporting their claim of no bias. Not only does this not account for the possibility of ice shattering, but aerosol measurements here appear to be missing data, since all aerosol measurements immediately vanish. I had hoped to confirm this, but the UTC times extend beyond those of the research flight they are showing, which appear should end around ~23 UTC. I urge the authors to revisit this figure and their arguments.

Perhaps I could have been clearer initially in requesting case studies. I had hoped the authors could show cases of level flight legs which have relatively similar aerosol concentrations within and right outside of the cloud – ideally where the cloud boundaries are observed at $RH_i \geq \sim 100$. In fact, aerosol concentrations would likely be slightly lower than outside of the cloud due to the activated INP. I am open to arguments of why such cases are not found and/or why this is not a plausible verification method.

Addressing figure r2:

The authors show a scatter plot of N_i vs D_i and argue that since there are multiple modes of $N_{aer500}>0$ within different ranges of the variables, that “This result indicates that there is no evidence showing that sublimation of small ice or shattering of large ice could be the main cause of large aerosol occurrences measured.” I do not understand why. There are indeed likely ranges where a particular bias may be

more likely observed, but without showing actual occurrence frequencies of data at given N_i and D_i it cannot be determined. The overlap of datapoints makes it impossible to determine whether $N_{aer500>0}$ occur at greater frequencies at large D_i relative to small D_i . In fact, the mode of $N_{aer500>0}$ at $1.5 < \log_{10}(N_i) < 4$ and $1.3 < \log_{10}(D_i) < 2.3$ could be argued to be associated with relatively larger IWC, where such biases are suspected to result from shattering.

Addressing Table r3:

The authors show that 33% of in-cloud samples have $N_{aer500>0}$, and state “the existence of ice particles does not always directly cause the occurrences of large aerosols.” It could possibly be argued that biases are minimal if most samples do not detect N_{aer500} . I mentioned above I wish they had applied more “physically relevant” in-cloud thresholds as requested, which would likely increase the frequency of $N_{aer500>0}$ in clouds.

Addressing Table r4:

The authors show that the ratio of $\text{in-cloud} + N_{aer500>0} / \text{clear-sky} + N_{aer500>0}$ is much greater when restricted to supersaturated conditions than when using all samples regardless of RH_i. They state “If ice sublimation or ice shattering is the main cause of large aerosols, one would not expect to see large aerosols occur more frequently when ISS is available.” I’m unsure of this, simply because we should expect the presence of cloud to be more likely at ISS. Therefore, we would expect the ratio at ISS to be greater for this reason. Perhaps I’m misunderstanding this point.

Addressing Figure S6&7:

While I’ve been especially skeptical of the aerosol measurements and the authors’ arguments in favor of them to this point, I like the clear-sky aerosol diagnosis methodology. Seeing as this method addresses the in-cloud biases, I recommend the authors use this clear-sky methodology throughout the manuscript—although I am open to continued debate on biases if the authors wish to provide additional evidence.

This also seems to improve the output in the Figures. Particularly for Figure 4, panels g-i show a sharp change in occurrence frequency trends from – to + N_{aer500} . This feature nearly vanishes in Figure S6.

Please give credit to D’Alessandro et al. (2023) who used a nearly identical methodology, especially considering there is overlap in authorship of both papers.

Major Concern 2:

The authors have addressed this concern by providing sensitivity tests of two figures, showing major trends are still observed. Assuming sensitivity tests applied to other analyses in the study reveal consistent results (e.g., the ML method, using the clear-sky aerosol method if the authors choose to do so), this point was adequately addressed.

Major Concern 3:

By updating the dataset, the authors show RHi as a single predictor variable is much more successful than Naer500, whereas previously both predictors had approximately similar predictor scores. The authors now also discuss including T into the predictions as a method of testing how the variables' predictor scores change when introducing a form of noise into the calculation. I find this unnecessary and do not necessarily agree with the logic, since there is no way to see how the noise is amplified in each case. I would avoid this latter point altogether.

The authors have adequately addressed my other concerns in this section.

Bibliography

D'Alessandro, J. J., G. M. Mcfarquhar, J. L. Stith, J. B. Jensen, Minghui. Diao, P. J. DeMott, and K. J. Sanchez, 2023: An evaluation of phase, aerosol-cloud interactions and microphysical properties of single- and multi-layer clouds over the Southern Ocean using in situ observations from SOCRATES. *Journal of Geophysical Research: Atmospheres*, <https://doi.org/10.1029/2023JD038610>.

McFarquhar, G. M., and Coauthors, 2017: Processing of Ice Cloud In Situ Data Collected by Bulk Water, Scattering, and Imaging Probes: Fundamentals, Uncertainties, and Efforts toward Consistency. *Meteorological Monographs*, **58**, 11.1-11.33, <https://doi.org/10.1175/amsmonographs-d-16-0007.1>.