We thank the editor for her interest in our paper and hope that our responses to her comments adequately address the concerns raised. We also note that her concerns regarding the above-cloud absorbing aerosol impacts on CER retrievals were also raised by Reviewer 2. We address those concerns in our response to that reviewer, and our responses here often refer to those.

General comments:

The introduction would benefit from a stronger tie-in to the cloud-aerosol interactions motivating ORACLES, currently it is highly general. Specific motivations include smoke-cloud microphysical interactions; the time evolution in CER towards discriminating the boundary-layer semi-direct effect from the microphysical interactions; the impact of intervening absorbing aerosol on the CER retrievals.

> We appreciate the editor's suggestion. That said, this paper has a narrow focus on the microphysical retrievals, and probe observations, that have relevance beyond ORACLES and the specific motivations for that campaign. As such, we'd prefer to keep the introduction tailored to that narrow focus and let the Redemann et al citation do the lifting on the details/motivations of ORACLES itself.

Overall I was surprised to not see more language about overlying smoke in the manuscript and its cases. Did I miss something? Was this not an issue?

> We thank the editor for raising this concern, which Reviewer 2 did also. While we did consider above-cloud aerosol effects on a regional/seasonal statistical basis, we regret that we did not consider these effects on the specific case studies shown here. Please see our response to Reviewer 2 for a detailed discussion of the impacts of above-cloud aerosols on the CER retrievals, an additional analysis we performed to account for these impacts in the case studies shown, and subsequent changes made to the text and figures.

Specific comments:

Paragraph spanning lines 56 to 69: why aren't SSA assumptions of the overlying smoke on the CER retrievals mentioned here?

> The discussion in this paragraph primarily is focused on cloud radiative model assumptions themselves. That said, in response to the Reviewer 2, we have added text on forward model errors due to ignoring above-cloud absorbing aerosol layers to Sections 4.1, 4.2, 5, and 6 that we think addresses the concern you raise here.

In the case of ORACLES the overlying smoke layer was also humid enough to support clouds at times near the top of the aerosol layer. The RSP also provided useful retrievals of these mid-level cloud properties, as shown in Adebiyi et al 2020. This paper would be worth citing, perhaps at the end of line 172 or as an example of CER retrievals helping to determine aerosol-cloud interactions in the ORACLES regime.

> We thank the editor for reminding us of this paper and have taken her suggestion to add the citation to the end of line 172 (now line 174), along with some minor text modifications.

Line 141: what is the lower limit of the drop size detected by CAS? Is it also 3 micron? It is also my understanding that the smallest dropsize bin of the PDI was too noisy to use, perhaps because of interference from the airplane's electrical system. Can the ORACLES PDI PI comment? For the thin, polluted clouds sampled during ORACLES, the lower size limits will be important. How did the Nds compare?

> While CAS can detect particles as small as 0.5µm diameter, for the ORACLES Microphysics dataset a lower limit of 3µm diameter was applied to avoid counting aerosols as cloud droplets. We note also that, for the merged DSDs, a lower limit of 50µm diameter is applied to 2D-S. More information on the probe size ranges/limits, as well as campaign-wide comparisons of probe Nd, LWC, etc., can be found in the O'Brien et al. readme file cited in the paper.

Regarding potential noise interference from the aircraft's electrical system on the smaller size bins of PDI, this was only an issue in the 2017 and 2018 deployments; PDI did not suffer this issue in the 2016 deployment data used in this study.

Section 2.4: continue to be puzzled why the SSA of the overlying smoke layer isn't mentioned as an additional uncertainty within the eMAS retrieval.

See our response to Reviewer 2 and to your comments above.

Top of p. 10: eMAS issues seem like they belong in data section.

We respectfully disagree, as this section is intended to highlight specifics of the eMAS operations during the ORACLES 2016 deployment, whereas the data section is intended to provide general details on the retrieval products and their heritage/theoretical basis.

Fig 1 caption: provide total # of flights in each category.

Good suggestion. Done.

Line 419: in combination, for LWP, the bias is about 10% (Grosvenor et al. 2018).

Indeed, biases in COT and CER will propagate to LWP calculations that, for bi-spectral imager retrievals, are proportional to the product of COT and CER.

Figs 7-9: What was the AOD overlying the cloud in both 20 Sept cases? Same question applies to the 14 Sept case later on.

The AOD overlying the cloud in both 20 September case studies, obtained from co-located HSRL-2 curtains, was just under 0.6, which we now indicate in the text; for the 14 September flight, HSRL-2 did not operate. As we discuss in our text additions on the impacts of this above-cloud aerosol layer, the specific impacts on the CER retrievals due to NIR absorption and solution space non-orthogonality act to increase CER, though to varying degree.

Fig. 10: are the cloud DSDs unimodal throughout the vertical column? Might be good to examine the actual DSDs near cloud top to understand the differences better. Maybe the issue is that the 2DS is having trouble picking up drops near the smaller end of its range, so that the DSD looks more bimodal within the CAS/2DS combined distribution. Is there any evidence of coincidence undercounting by the CAS? It would also be good to show the vertical profiles in LWC, allowing you to add in the bulk King-derived LWC. How thermodynamically well-mixed was the cloud layer?

While we share the editor's interest in understanding why the cloud probes differ, this paper is focused primarily on the remote sensing retrievals of CER and the impacts of bi-spectral retrieval assumptions on CER agreement (or disagreement). As such, we think that attempting to diagnose differences in the probe datasets is well beyond the scope of this paper. We simply take the probe data as they are provided in the public ORACLES archive. We do note, though, that analyses that include the campaign-wide LWC profiles from the probes in question are shown in the O'Brien et al document, cited in the paper, describing the ORACLES Microphysics datasets. Furthermore, we do agree that this subject perhaps deserves a dedicated study, and have added a statement to this effect to the discussion in Section 5 (lines 1058-1060).

Table 2: why are the polarimetric optical depths so low?

These are vertically weighted COTs, or essentially the optical depth within the cloud where each retrieval has its peak sensitivity. Since the polarization signal is a single-scattering phenomenon, its sensitivity lies at the very top of the cloud, as shown by its corresponding weighting function in Fig. 10 (c). The spectral retrievals, on the other hand, have peak sensitivity much deeper in the cloud, a fact that drives the discussion on the impacts of vertical heterogeneity.

p. 41: how confident are you in the CO2, CH4, H2O above-cloud optical depths/loadings? If I understand correctly these are coming from a reanalysis - is that MERRA2? You mention doubling the concentrations, these might bring you closer to the actual values. How do the reanalysis H20 values compare to those measured on the aircraft? This error source should be genuinely quantifiable using the ORACLES H2O in-situ dataset in contrast to the statement made on line 909. THis may be addressed in Pistone et al 2024.

> This is an excellent question. The short answer is that we are not confident in these column loadings. For the eMAS retrievals (and their spaceborne MODIS and VIIRS counterparts), we use NCEP GDAS reanalysis profiles for H2O, and standard atmospheric profiles for the remaining gases. This is an obvious source of error in our atmospheric corrections that was the rationale for investigating the sensitivity to these assumptions and their impacts on the CER retrievals. That said, rather than attempting to estimate the exact biases in these reanalysis column loadings, we instead chose to estimate the impacts of extreme biases, here 100% errors that we assume are larger than the actual biases. Even under these extreme error assumptions, the impacts of atmospheric correction errors cannot explain the retrieval differences found in the case studies.

p. 44: more effort can and should be made to understand the differences between the two probes. How about picking a long level incloud-leg, ideally near cloud top, and comparing an average of the full DSD. Is there evidence of bimodality? How does Nd compare? What is the flow volume of the 2 probes? Is the sensitivity to the smallest drop sizes the same?

> Again, while we share the editor's interest in understanding the differences between the probes, we think attempting to answer these difficult questions here is well beyond the scope of this paper. We do agree, though, that such an investigation is of interest to the community, particularly the users of the ORACLES datasets, and should be pursued separately, and have added a statement to this effect in Section 5.

Section 5: why isn't the aerosol overlying the cloud mentioned for these cases? Could it be neglected? Did you select 2 cases with no overlying AOD? I don't recall reading that and wonder if I missed it.

> See our responses above and to Reviewer 2 that detail additional analyses and text that we think now address this concern.

Section 6: Stronger guidance should be provided here to help potential users of the publicly-available eMAS cloud property retrievals. Which one would you recommend?

> This is a good suggestion. We have added a paragraph at the end of Section 5 that provides advice to users.