Response to reviewers

The authors thank both reviewers and provide responses to their comments in blue text below:

<u>**REVIEWER 1**</u>

First review of "A method to retrieve mixed phase cloud vertical structure from airborne lidar" by Crosbie et al.,

January 15, 2025

Reviewer Recommendation:

Somewhere between Minor and Major revisions required

Summary:

The submitted manuscript introduces an empirical model in an attempt to separate the origin of depolarization observed with a lidar. In my understanding, the authors take observations that are expected to be from liquid only clouds, as screened by temperature, and tune an empirical model to tightly match the depolarization profile observed. This model accounts for depolarization due to multiple scattering. When taking the same model and applying it to clouds that are not necessarily expected to be liquid only, deviations are then attributed to ice. In general, this approach seems reasonable and caveats and assumptions are fairly well documented.

In a world that is often filled with algorithms that over promise and under deliver, this manuscript is, in my opinion, candid with limitations and expectations. I don't often get to write that and I think it is important to note. In general, the paper is well written; it has generally excellent clarity and citations are appropriate in my opinion. It fits well within my understanding of the scope of AMT and I believe it is a contribution that should be published.

I note some elements that I find unclear, and I believe some revisions will improve the draft. That said, it is my recommendation that this manuscript be accepted with revisions (somewhere between major and minor).

Major Comments:

1. My prime concern is that I am unclear how far to trust the validation of this technique. In general, I am sympathetic to the fact that you are using airborne measurements and that they are generally infrequent (especially because in this case you require 2 planes in the same patch of sky simultaneously) and you are trying to see into a region where few technique exist. However, in Figure 6 & 7 the flight track of the Falcon and your measurements seem to be worryingly far away for a worryingly large fraction of time. This seems to be the crux of the problem: lidar has poor penetration depth into clouds but you need coincident measurements to validate the technique. My specific comments/questions are:

   The reviewer raises concerns that are unfortunately a central theme of airborne observations, particularly for dynamically evolving clouds. Despite the best efforts of the ACTIVATE team to coordinate the aircraft, some degree of mismatch in sample volume between the two aircraft (whether in 3-D space or in time) was almost always unavoidable. We have approached the problem in the best way known to us, including additional corrections (such as approximating the contribution for wind drift) that have not been routinely undertaken in other studies involving the ACTIVATE aircraft matchups.

   It should also be added that as it relates to airborne campaigns, ACTIVATE may offer perhaps one of the best datasets as it relates to spatially coordinating two aircraft with over 70% of the nearly combined 1200 flight hours with two aircraft involving them being within 6 km and 5 min of each other; beyond this point we should note that we were careful in this study to use data that involved sufficiently close spatiotemporal coordination to do this analysis robustly. Thus, while not perfect, the spirit of advancing science and technology should use datasets such as this one that are as good as available and meets the standards of the investigating team with decades of experience in this line of work.

   1. It might be best to specify how much total data overlap there is (in terms of hours or kilometers or measurements or something else), i.e. when the Falcon is within the realm of your measurements that you trust (to my novice eye, 15-20 km range on Figure 6 seems to be reasonable overlap but there seems to be almost none from 30-40 km on the same figure).

      We have added an overlap duration metric as an additional column in Table 3 and we thank the reviewer for this suggestion. For additional clarification, the quantitative evaluation discussed in Section 4 (and shown in Figure 10) would not include almost all of the 30-40 km region that the reviewer mentions, because it does not meet the overlap

requirement (illustrated in Figure 9). The examples described in Section 3-5-3.7 and shown in Figures 6-8 were picked to highlight a diversity of cloud types. In the center part of Figure 6, the qualitative agreement in the pattern seen near cloud top with the lidar compared to the deeper in situ is worth stating, but of course we agree that it should not be used as validation. We have made an adjustment to the text in Section 3.5 to clarify this further.

In Section 3.6, (the reviewer mentioned Figure 7 in the introductory remarks for this comment) we have already provided a caveat to any regions that were not vertically collocated but we feel that our commentary/interpretation of the cloud scene is appropriate and adds to a contextual understanding of the overall cloud processes. Again, any regions that do not satisfy the screening criteria are not included in the quantitative validation in Section 4.

2. Lines 428-429: It is clear to me that you are picking +-300 meters as a threshold because you need to have some reasonable sample size. However, it is unclear to me how sensitive to this limit (+- 300 meters) your results are? 300-meter penetration into liquid clouds seems optimistic. Your Figure 3 & 4 suggests that your penetration depth into clouds may reach this threshold but can also be limited to 50-60 meters. As I read this, I think either: 1) your results need to be shown to be robust for various depths of observation or 2) the sensitivity to that change needs to be established. At minimum, if you use a more restrictive threshold, I believe you need to comment on what it does to your interpretation of results.

The Δz sensitivity was investigated but was not ultimately included, mainly for brevity. We include the sensitivity analysis below. As the reviewer notes in the first comment above, it is essential to establish a vertical threshold, but the best choice of value is uncertain. Smaller Δz results in more conservative screening but that does not necessarily translate into a fairer validation because the sample size may become unworkably small, and it amplifies biases related to subjective altitude choices for the Falcon.

While the penetration depth is often less than 300 m (as noted), the local variability in the cloud top is often of a similar magnitude, at least in some of the cloud scenes. Along a similar argument as made for using a larger Δx to minimize residual collocation errors, the larger Δz also helps to

suppress collocation error. If the collocation error could be completely eradicated allowing record-by-record comparisons ($\Delta x \sim 50$ m), then one could expect $\Delta z < 10$ m.

Figure 1 shows how the MAE for the mix classification with $\mu_{thresh}=0.14$ changes with Δz, where $\pm\Delta z$ is the vertical depth threshold. The MAE remains nearly constant through approximately 160 m then drops at lower Δz. Figure 2 shows how MAE varies with μ_{thresh} for the mix classification. The behavior is generally quite consistent although the location of minimum varies in the 0.08-0.14 range. For $\Delta z = 200$-500 m, similar MAE performance could be claimed by using μ_{thresh} anywhere with that 0.08-0.14 range. Figure 3 shows the number of subsegments (defined by Δx and Δz) that are suitable for making comparisons between lidar and in situ. A subsegment needs enough overlapping data points ("pixels") and sufficient cloudy data from the in situ sampling to be considered suitable. As Δz drops below 200 m, the number of subsegments drops considerably. The mean number of "pixels" per subsection shown on the right axis of Figure 3 also shows a decrease because the capture area Δx by Δz decreases.

Our conclusion is that between 200-500 m the choice of Δz is insensitive and the result is consistent. Using a smaller Δz gives the illusion of better performance (under the metrics for performance we have used) but we believe that it is driven by the precipitous loss of qualifying data points rather than an actual improvement. It is possible that this could be explored further with a future cloud dataset that has more uniform cloud top heights and less concern for collocation error.

We have added the following to the description in L433:

"The analysis was found to be insensitive to vertical zone size in the 200-500 m range."

Figure 1

Figure 2

Figure 3
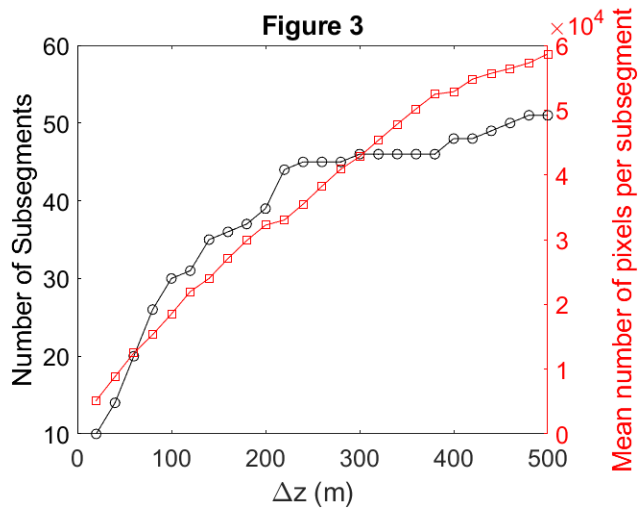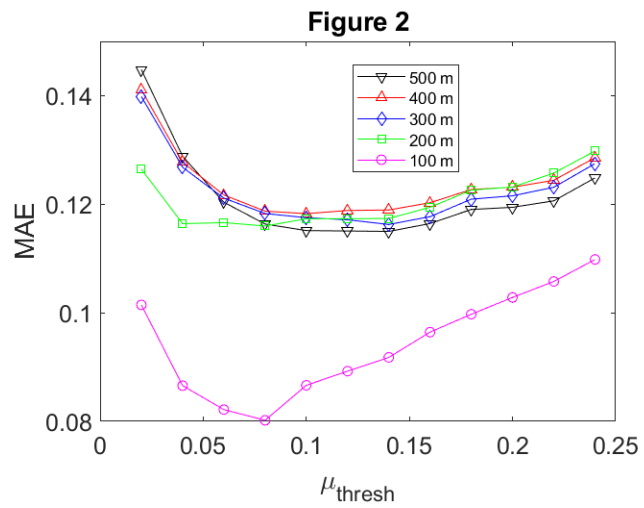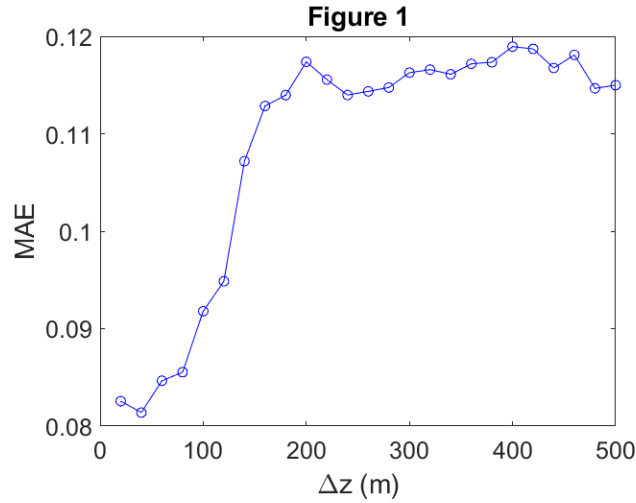
3. Line 380-382: Given that you are using these examples as validation for your method, it seems like a non sequitur to specify that the method

works and that the Falcon's data can't be used because it was too far away.

This comment is addressed in the responses to 1 and 2 above. To reiterate, segments of the data that do not meet the collocation requirements are screened out from the validation in Section 4. We still feel it appropriate to discuss the time series and show examples like these because it provides context for a reader. We repeatedly caveat any interpretation when there is not direct collocation.

4. I presume the GRD case from Table 1 has no flight validation data. Is that true? If that is true…is it a fair control case? If not, you should more clearly say how that data is compared/validated. For example..where is your temperature data coming from?

None of the water-only control cases introduced in Section 2.2.1 overlap with the King Air – Falcon matchups from ACTIVATE. So, in that sense, none of the control cases are validated with in situ data. The control cases were picked from a range of conditions that crucially spanned as much of the range-to-cloud extent as possible. The temperature was determined from radiosondes or dropsondes.

2. Section 2.3.1 and Lines 412-414: My understanding of the observation of oriented ice crystals is that their properties are highly dependent on the angle of observation. For example looking at Noel and Chepfer 2010, it is my understanding that a change from 0.3 degrees from nadir and 3 degrees from nadir basically removed CALIOP's sensitivity to oriented ice crystals. With that being said, are your observational legs done at nadir exactly or is there some angle of attack? It the plane pitching/rolling in any way that would change your sensitivity? Does that affect your measurements of oriented ice crystals? A comment clarifying your sensitivity would be warranted in my opinion.

Noel, V., and H. Chepfer (2010), A global view of horizontally oriented crystals in ice clouds from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO), J. Geophys. Res., 115, D00H23, doi:10.1029/2009JD012365

There is some variability across different platforms, but the King Air typically flew with a 2-4 degree nose up pitch angle and the lidar was registered within 0.5 degrees of the fuselage datum. Note, for clarification, it is the nadir offset angle of the lidar relative to vertical that is the critical angle (the angle-of-attack mentioned by the reviewer is the angle of the airframe relative to the airstream and therefore incorporates climb/descent angles).

The maneuvering of the aircraft (pitch/roll) will certainly affect the sensitivity to signatures of oriented ice, as the reviewer correctly notes. However, it is not clear how to provide a statement/comment on this sensitivity, other than to state that the aircraft is oriented off-nadir. We have added this to Section 2.3.1. Even though the lidar is off-nadir (typically at a comparable angle to CALIOP), the ability to identify regions that may be "polluted" by the complicating influence of HOI was still deemed to be important.

As described in L307-308 we did not expect a strong influence from oriented ice in the ACTIVATE matchup cases because most of the mixed phase cloud occurrences were during post-frontal/cold air outbreak conditions. These conditions are turbulent, high LWC environments and therefore promote riming that is assumed to further suppress the HOI signature.

The comments in L412-414 are intended as a cautionary statement and to potentially stimulate further evaluation of the angular dependence of specular reflections. It is possible that with more detailed vertically-resolved information (compared to layer-integrated approaches) the conclusions about the prevalence of oriented ice (and mixing relationships) could evolve. However, we feel that further analysis is out of scope.

Minor Comments:

1. Lines 122-124: You mention cross talk in the molecular channel due to imperfect extinction by your HSRL filter. This physical effect is absent your equation 2. I would suggest including that term.

    No, because the series Eq 1-13 explain how we manipulate the signals for this method. We do not use the SR, via Eq (4), for quantitative evaluation of the cloud extensive properties. The only place that SR is used is to establish a reference cloud top altitude. At SR=50 the cross talk is assumed negligible.

2. Equations 10/11/13: It is unclear to use the same variable (z) as both the integrand and the upper bound of integration. Suggest integrating over z' or some other stylized version to differentiate your two variables.

    We added text to clarify that the depth coordinate z starts at cloud top and extends down into the cloud. The addition of other stylized variants on dz was deemed to be more confusing especially for Equation 13 where we already have the superscript ∥. The prime has already been used to denote differentiation with respect to depth in Equations 17 and 18. An asterisk has already been used for Equations 14-16. We feel that the text clarification is sufficient.

Line 192: I would expect this value to be lower (like maybe 8 to 10). I think therefore that a citation for the lidar ratio used here is needed.

References added. Certainly, by conventional definitions, 18-20 is typical.

3. Table 1: What is "Cloud Temperature"? Is it really cloud-top temperature or average temperature or something else? Please clarify.

It is an estimate of the temperature where the measurement was made. Cloud top temperature would be appropriate except for the zenith case (GRD), where we do not necessarily know where cloud top is. The point of listing these temperatures is to demonstrate that they are not frozen. Given that it is not really the main point, it started to get laborious to further document additional details.

4. Table 2: I would suggest adding the RMSE value that would be calculated if using your "Final" coefficient values as another row. It would be helpful to see how much your averaged coefficients alter the fit to each test case.

Added, thank you.

5. Table 2: Why is your GRD case included in Table 1 and not here?

While it was desirable to have an additional data point at close range to confirm the near-field multiple scattering shown in Figure 2, it did not seem appropriate to train the MSD model on zenith data. While it would be interesting to assess any differences in the accumulation of depolarization from the (typically) more gradual increase in extinction found at cloud base, it was deemed less relevant to the intended end goal.

6. Figure 2: I would suggest changing your colors for each case to perhaps use different symbols (diamonds, triangles and so forth). For example, I am having a hard time differentiating ACTHIGH and GRD. Additionally, I am not Red-Green colorblind but CPEX and ACTLOW1 might be difficult to differentiate for someone that is.

We have implemented the different symbols suggested by the reviewer and this should circumvent any ambiguity caused by color. Thank you for this suggestion.

7. Figure 3: Why is there a huge gap in measurements from 60 meters to 120 meters? Is that last data point truly trustworthy or are your data filters kicking in with just one point serendipitously falling into the end?

From L217-219: "For implementation of the method, Δz is allowed to vary to keep increments in γ∥ constant (up to the native acquisition of the measurement Δz = 1.25 m) and this has the advantage of mitigating the decrease in signal-to-noise ratio with depth in the cloud that affects both α* and δ."

The lines in Figure 3a,d,g are the "raw" signal while the dots are the re-gridded values. This clarification has been added to the figure caption to avoid any ambiguity. Thank you for mentioning it.

8. Figure 3: It might be helpful to add an ordinate axis for optical depth instead of just using physical depth.

   We tried this but it was more clumsy than beneficial.

9. Figure 3 & 4: I would include in the caption what the difference between gray and blue does is (specifically looking at panels C/F/I). I presume it is for observed and modeled depolarization, respectively, but it would be good to clarify.

   Agreed, thank you.

The manuscript presents an empirical methodology for identifying ice and liquid water features within atmospheric clouds. Hydrometeor characterization is a well-documented challenge in cloud research and has been addressed in prior studies. However, this work is the first to develop and test a method specifically designed for an airborne lidar system. The study leverages extensive measurement campaigns, incorporating lidar observations and in situ data, adding significant value to the analysis.

While the manuscript outlines the general approach effectively, certain methodological details require further elaboration. Key geometrical considerations, such as the field-of-view and distance to the cloud, were addressed, and the results support the reliability of the approach. Nevertheless, questions remain regarding the quality of the lidar data, particularly the reported depolarization values, which appear anomalously high.

We would lean on the reviewer to provide more quantitative support for their statement that the depolarization values appear high. We would also like to reiterate that if the reviewer is used to layer accumulated depolarization ratios for clouds, then the maximum range-resolved depolarization ratios will be higher (typically a factor of ~2). In addition, if the reviewer is used to very narrow FOV systems at close range then the multiple scattering depolarization will be much smaller in those cases.

We further refer the reviewer to Figure 2a of the manuscript. Here the layer integrated depolarization ratio (D) is compared to the layer integrated backscatter (γ). At different ranges that modulates the amount of multiple scattering, the relationship conforms to the well-established Hu parameterization of water cloud integrated depolarization ratio, for a typical cloud lidar ratio. If the depolarization ratio were anomalously high, this agreement would not hold.

The integrated depolarization (Eq 13) of the CPEX case shown in Figure 3c is 0.197. For reference, CALIOP sees layer depolarization sometimes more than 0.3 for water clouds (Hu 2007).

Hu, Y. (2007). Depolarization ratio–effective lidar ratio relation: Theoretical basis for space lidar cloud phase discrimination. Geophysical research letters, 34(11).

Additionally, the manuscript suggests that the system is configured to measure only in a downward direction. This setup has significant implications and needs to be explicitly addressed. Unlike ground-based lidars, the downward-facing configuration could lead to depolarization increases due to multiple scattering, which may at some point merge

with the single-scattering depolarization of ice crystals. This is contrary to the typical behavior observed in ground-based systems. Moreover, strong attenuation of the signal in liquid layers near the cloud top might obscure ice or liquid water signatures beneath these layers.

To clarify, these airborne systems are operated in a near-nadir configuration while on the aircraft. However, they can be operated in a near-zenith configuration in the laboratory. This fact was stated in the manuscript (L183-185) in the description of the ground-based control case. The manuscript does not suggest that the system is configured to measure only in a downward direction, as the reviewer asserts.

It is challenging to parse the comments in this paragraph from the prior comments that claimed the depolarization was too high for our system. Specifically, it appears that the reviewer is acknowledging that cloud top nadir views may result in more depolarization than they are used to with ground systems. The fact that parts of the scene are obscured because of attenuation is a concern of both upward and downward looking scenarios.

Finally worth noting: yes, absolutely, at some point water multiple scattering depolarization will be equivalent or greater to the typical signature from ice crystals. For some ground-based systems this may never be an issue and therefore could significantly simplify the problem.

To enhance the robustness of the method, synthetic scenes or ground-based observations might be beneficial in the future.

An excellent suggestion. We hope that data from other systems could be used to train a MSD empirical model using suitable water-only training sets and would welcome comparisons.

Specific Comments:

- Line 100: Using the HSRL, backscatter and extinction coefficients can be unambiguously derived. However, how are these parameters affected in liquid clouds? Could crosstalk introduce significant errors in the measurements? Additionally, why was the extinction from the molecular channel not considered in this approach?

  As stated in L122-125, there is a possibility of cross talk at very high SR. This is one of the reasons why the HSRL molecular signal is only used to "calibrate" just above cloud top. There may be additional information contained within

the molecular signal within the cloud but we are not using it at the current time. This is clearly stated in Section 2.1.5.

- Equations 1, 3, and 5: These equations assume an idealized polarization system. How do the authors ensure the absence of polarizing effects, given the potential inefficiencies in each channel? (Mattis et al. 2009, Freudenthaler 2016). Furthermore, the reported depolarization values for liquid layers appear unusually high.

  See earlier comment regarding the depolarization ratios. The waveplate and channel gain ratios are calibrated at least once per flight and these ratios have been found to be very consistent. See Hair et al. 2008 for details (this is referenced in the manuscript).

- Equations 1 and 2: The use of the same constant for both channels requires justification. Since the detectors and optical paths differ, how is it ensured that the constants are identical? And shouldn't the cross-polarized channel in Equation 3 have a different constant? A polarization calibration might be necessary.

  There are different constants. Their ratios are G_m and G_{dep}, respectively. The equations already have this included in the determination of the X terms.

- Equations 8 and 9: These equations describe the attenuated backscatter from the system to range z, not from the cloud base to z. The left-hand equality seems incorrect as written. Shouldn't the left-hand side be divided by ($T\_norm^2 * T\_m^2$)?

  By cloud base, we assume the reviewer means cloud top (in the context of the way that Section 2.1.5 is explained).

  No. The intent is to write an attenuated backscatter with respect to a reference at cloud top, so the left equality describes what we mean it to mean: cloud attenuated total (molecular + particle) backscatter for co-and cross-polarized components.

- Equations 10–13: These equations do not appear essential for introducing integrated quantities. Please consider removing them to streamline the manuscript.

  Respectfully, we disagree. Splitting the y terms (10-11) in this way is a new approach and so we want to define it clearly. The purpose of Eq 12-13 is to

remove any lingering ambiguity about using range-resolved versus layer accumulated depolarization profiles.

- Line 198: The explanation regarding larger lidar ratios for ice crystals underestimating extinction more than liquid clouds is unclear. While multiple scattering contributes to extinction underestimation, this applies to both cases, ice or liquid. Please clarify. What reference supports Equation 16, and what are its limitations? The equation seems to assume a constant extinction coefficient profile, so additional reasoning is required.

  It is not obvious what needs to be clarified in L198. If the lidar ratio is greater than the estimate, the true extinction will be more than the estimate.

  The form of Equation 16 is a modification of the method of Roy and Cao (2010) which is referenced in this section. The limitations are specified throughout Section 2.2.2. It is also made clear that this is an extinction profile and therefore a function of z (not a constant as the reviewer suggests). The lidar ratio is a constant.

- Equation 19: With five constants to retrieve, how good is the performance of this retrieval? Given the potential for multiple solutions, how do the authors ensure an optimal and unique solution?

  For clarification, not all are constants: r_2 is a function of range-to-cloud. The quality of the parameters is assessed with respect to the RSME shown in Table 2. The reviewer is justified in asking about the potential for additional suitable parameters. There were correlations amongst the parameters that resulted in zones within the parameter space with similar performance and caused slow convergence. After mapping these dependences, a "brute force" iteration was employed to minimize the risk of a better solution somewhere else.

  Our intention is not to claim that the form of Eq19 is the best representation of multiple scattering. Indeed, there could be physically based or machine learning approaches that are superior. The potential for multiple solutions in the current set of parameters is not overly concerning because we are not trying to interpret their meaning in a physical sense. The uncertainty generated by the RMSE in Table 2 or the bounds in Figure 5, would be good to improve upon but the current levels do not preclude its usage.

- Figure 3: The differences among the three cases are difficult to discern. Extending the vertical range in the figure could enhance interpretability.

The range shown currently spans the entire range of workable data. In the first (a,b,c) and second (d,e,f) cases the retrieval would not be performed any deeper because we are already at the extinction limit. In the third case (g,h,i), which is the translucent case, we have already exited the cloud and there is nothing more useful.

- Section 3.3: The utility of the retrieved microphysical information is unclear. The primary instrument, the 2DS optical array, only detects hydrometeors larger than 90 μm. This limits its ability to observe cloud droplets, which are typically smaller than this threshold.

  The reviewer has missed some critical aspects in their reading of this section. First, as listed in the first sentence, both the FCDP and 2DS are used to interpret the microphysics. The FCDP provides particle distributions from 3-50 μm (stated in L345). The 2DS can detect particles smaller than 90 μm; however, it cannot distinguish between liquid and ice (based on sphericity) below 90 μm (as stated in L347). We have no further response to the reviewer's comment.

- Sections 3.5–3.7: When introducing the retrieval scheme, it would be beneficial to first present the lidar-observed variables (e.g., attenuated backscatter, depolarization).

  Perhaps we could include lidar curtains, but there is already quite a large amount of information held in Figures 6-8 and at this stage in the manuscript we wish to focus on the comparisons between the derived phase and the in-situ data.

- Figure 5: Depolarization values up to 50% in warm liquid clouds seem unusually high. Such values are typically associated with ice crystals. What depolarization values are observed for ice crystals?

  Please refer to the earlier response to the alleged high bias in depolarization. The ice threshold was set at 0.35 (L295).

  Therefore, deep into dense clouds it would be ambiguous, if not impossible, to isolate an ice signature. This is not a fact that we have ignored in the analysis and development of the method. In the validation section, the conclusion that an ice extinction fraction of 14% defines the onset of the MIX classification is essentially stating that smaller amounts of ice are, on average, not possible to isolate from the signature of multiple scattering from water. With more aircraft matchups we may be able to stratify the data by optical depth (or

some other measure) to improve the interpretation of the threshold. Currently, the datasets are too sparse.

- Finally, the manuscript would benefit from a detailed author contribution statement. Given the technical complexity of the work and the number of coauthors, such a statement would clarify the specific roles of each contributor.

The author contribution statement is accurate as written.

References:

Mattis, I., Tesche, M., Grein, M., Freudenthaler, V., and Müller, D.: Systematic error of lidar profiles caused by a polarization dependent receiver transmission: quantification and error correction scheme, Appl. Optics, 48, 2742–2751, 2009.

Freudenthaler, V.: About the effects of polarising optics on lidar signals and the Δ90 calibration, Atmos. Meas. Tech., 9, 4181–4255, https://doi.org/10.5194/amt-9-4181-2016, 2016.