Please find my comments in red after the author's responses.

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).

The reviewer acknowledges that you consider height-resolved values rather than the integrated depolarization.

Depolarization is closely linked to the multiple scattering effect—generally, the greater the contribution of multiple scattering, the stronger the depolarization. The relationship from Hu et al. (2006), as you mention, is a well-established method for relating integrated depolarization to the single scattering fraction in the integrated attenuated backscatter. As you also note in the manuscript, this relationship is largely independent of geometrical features. However, the extent of multiple scattering, and consequently depolarization, does indeed depend on the lidar's geometrical configuration, the distance to the target, and its optical density.

Regarding Figure 2a: You state that the X-axis represents the integrated (attenuated) backscatter, whereas the caption describes it as the backscatter enhancement. Which description is accurate? I assume it refers to the total integrated attenuated backscatter (or integrated parallel attenuated backscatter?).

Hu et al. (2007) found that the (total) integrated backscatter follows this relationship. $\gamma = \gamma_{ss} * \left(\frac{1+D}{1-D}\right)^2$, where $\gamma_{ss} = (1 - T^2)/(2S_c)$. Does the pink line represent this relationship? Please elaborate on this explanation to make it clearer.

Furthermore, the variability range shown in Hu et al. (2007) cover approximately 0.1 - 0.3 for the integrated depolarization D and between 0.05 – 0.1 sr-1 for the attenuated backscatter γ. Fig 2a in the manuscript shows a range of 0.05 – 0.2 for the integrated depolarization and 0.03 - 0.06 sr-1 for the attenuated backscatter. For an integrated attenuated backscatter of 0.06 Hu et al 2007 reported values of D roughly between 0.13 and 0.17 while you report a mean value of 0.2.

In any case, finding values within a similar range to those of the CALIOP lidar does not necessarily validate the accuracy of the values reported in the manuscript, given the complex nature of multiple scattering in liquid-water clouds. As mentioned, the measured depolarization varies from system to system and depends not only on the cloud's optical thickness but also on factors such as the laser divergence, the receiver's field of view (FOV), the laser and receiver diameters, and the distance to the cloud target. In the case of the CALIOP lidar, multiple scattering effects can play a major role due to its large footprint while orbiting at approximately 705 km altitude.

Modelling studies have reported maximum depolarization values of 0.25–0.35 for a penetration depth of 100–150 meters in relatively opaque clouds (Donovan et al., 2015; Jimenez et al., 2020; Ahmad et al., 2022). However, even these studies consider systems with larger FOVs and small laser divergences (compared to the FOV). When laser divergence and FOV are similar—as appears to be the case for the HSRL-2 and HALO systems—the multiple scattering effect is expected to be weakest (Wang et al., 2021).

In Figure 3c, for instance, depolarization reaches 0.5 at a penetration depth of just 80 meters, which seems surprisingly high.

In my view, whether the depolarization values are overestimated or not does not affect the validity of the proposed methodology. However, I encourage the authors to acknowledge these potential polarization effects. The proposed approach for simulating the depolarization of water clouds (Table 2) would only be applicable to the lidar systems whose data were used for the training.

Hu, Y., Liu, Z., Winker, D., Vaughan, M., Noel, V., Bissonnette, L., Roy, G., and McGill, M.: A simple relation between lidar multiple scattering and depolarization for water clouds, Opt. Lett., 31, 1809–1811, https://doi.org/10.1364/OL.31.001809, 2006.

Hu, Y., Vaughan, M., Liu, Z., Lin, B., Yang, P., Flittner, D., Hunt, B., Kuehn, R., Huang, J., Wu, D., Rodier, S., Powell, K., Trepte, C., and Winker, D.: The depolarization – attenuated backscatter relation: CALIPSO lidar measurements vs. theory, Opt. Express, 15, 5327–5332, https://doi.org/10.1364/OE.15.005327, 2007.

Donovan, D. P., Klein Baltink, H., Henzing, J. S., de Roode, S. R., and Siebesma, A. P.: A depolarisation lidar-based method for the determination of liquid-cloud microphysical properties, Atmos. Meas. Tech., 8, 237–266, https://doi.org/10.5194/amt-8-237-2015, 2015.

Jimenez, C., Ansmann, A., Engelmann, R., Donovan, D., Malinka, A., Schmidt, J., Seifert, P., and Wandinger, U.: The dual-field-of-view polarization lidar technique: a new concept in monitoring aerosol effects in liquid-water clouds – theoretical framework, Atmos. Chem. Phys., 20, 15247–15263, https://doi.org/10.5194/acp-20-15247-2020, 2020.

Ahmad, W.; Zhang, K.; Tong, Y.; Xiao, D.; Wu, L.; Liu, D. Water Cloud Detection with Circular Polarization Lidar: A Semianalytic Monte Carlo Simulation Approach. Sensors, 22, 1679. https://doi.org/10.3390/s22041679, 2022.

Wang, Z., Zhang, J., Gao, H.:Impacts of laser beam divergence on lidar multiple scattering polarization returns from water clouds,
Journal of Quantitative Spectroscopy and Radiative Transfer, Volume 268, 107618, https://doi.org/10.1016/j.jqsrt.2021.107618, 2021.

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).

The comment refers to potential polarization effects that are unrelated to the gain ratios. These effects may be induced by optical elements within the system, including the beam expander, telescope, lenses, beam splitters, and polarization optics. I have learned that efforts have been made to account for these effects in the case of the HSRL-2 lidar (e.g., Burton et al., 2015).

Burton, S. P., Hair, J. W., Kahnert, M., Ferrare, R. A., Hostetler, C. A., Cook, A. L., Harper, D. B., Berkoff, T. A., Seaman, S. T., Collins, J. E., Fenn, M. A., and Rogers, R. R.: Observations of the spectral dependence of linear particle depolarization ratio of aerosols using NASA Langley airborne High Spectral Resolution Lidar, Atmos. Chem. Phys., 15, 13453–13473, https://doi.org/10.5194/acp-15-13453-2015, 2015.

• 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.

Thanks for clarifying this. It was not immediately obvious to me... So, c is the constant for the parallel channel, $G_m c$ for the molecular channel and $G_{dep} c$ for the cross channel.

- 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.

When I read cloud-attenuated backscatter, I interpret it as the attenuated backscatter within the cloud, which you refer to as the corrected signal. I suggest emphasizing this in the text by explicitly stating that you define attenuated backscatter as the corrected signal divided by the total transmission from the system to the cloud top (or cloud base if zenith-pointing).

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 γ 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. Ok!

- 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.
Yes, of course, the lidar ratio is set as a constant and not the extinction. I understand that. What is not entirely clear to the modification of the method from Roy and Cao (2010) leading to your Equation 16.

 The approach from Roy and Cao (2010) allows for the retrieval of the true extinction coefficient by correcting for multiple scattering using the Hu relation. In contrast, your approach accounts for multiple scattering in a more rudimentary manner by using the ratio of integrated attenuated backscatter to the maximum value (or the simulated γ_{rtc}). I can understand that you do not necessarily need to perform a sophisticated retrieval of the true extinction coefficient correcting the multiple scattering effect, instead you need α^* for your training and later identification, but equations 14-16 do not look obvious to me.

The term α^* represents the apparent extinction coefficient that produces the measured attenuated signal. To my understanding, the retrieved backscatter (and consequently, the retrieved extinction) will be lower than the true backscatter, regardless of whether the cloud is composed of liquid or ice. This is due to the apparent reduction in attenuation caused by the multiple scattering effect.

In the case of ice clouds, the extinction may be even more underestimated because the backscatter is multiplied by an incorrect lidar ratio to derive extinction from backscatter. This seems logical to me when considering the Klett-Fernald retrieval; however, it is not entirely clear to me in the context of Equations 14–16.

You write in Line 198: *As an example, α* would underestimate the true extinction profile for conditions where Sc>Sc,ref that may occur with sufficient ice content; however, as will become more apparent, the underestimate usually benefits rather than hinders ice 200 discrimination.*

I may have just missed the part where the benefits become more apparent.

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.

I was indeed referring to the ability to distinguish between droplets and ice crystals. The 2DS is unable to identify particles between 50 and 90 µm in diameter. If the FCDP detects particles smaller than 50 µm, you classify the pixel as a droplet, even though some ice crystals may be present.

Given these limitations and the challenges in finding suitable cases, as mentioned in Section 2.4, it was difficult to assess and evaluate the validation approach. Introducing supercooled liquid, drizzle, and droplets at that stage of the manuscript seemed unnecessary. After all, your empirical approach is focused solely on identifying ice and liquid water.

•   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.

I realize I may not have expressed myself clearly. I was referring to the range of depolarization values you observed in ice crystals. While I understand the difficulty of isolating the ice signature from the multiple scattering effects of water at low ice extinction fractions, what range of values do you typically observe in clear ice crystal cases? This information would also be relevant for the manuscript, alongside the the values for liquid clouds (which are at least provided for the reference pure liquid cases in fig 5).

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.

Line 443 refers to satisfy Equation 22. It should write Equation 21.