Dear authors,

Thank you for responding thoroughly to the reviewers' comments. After considering your exchange with them, I believe the manuscript is now ready for publication. It should be noted that the authors have correctly left some questions open, highlighting possible future developments.

However, I ask the authors to make only one change to a specific point raised by reviewer #2, which, in my opinion, has not been given due consideration.
Reviewer #2 notes that “There should be clarifying text to clarify the difference between deficiencies in lidar sampling that would result in errors, vs adequate sampling strategy (fast and high resolution) but averaging over multiple shots to get adequate SNR for a good DIAL or cloud property retrieval.”

The authors rightly note that the work of Gros et al. AMT 2014 is central to the focus of the discussion but do not fully engage in the clarification of how to generally treat DIAL under cloudy conditions, as per reviewer request. It would be appreciated if the authors could amend the relevant paragraph in section 2.2 with a more comprehensive discussion, following the reviewer's suggestion.

Best regards
Luca Lelli

From the second reviewer:
The comment about integrating over heterogneous backscatter is noted. There should be clarifying text to clarify the difference between deficiencies in lidar sampling that would result in errors, vs adequate sampling strategy (fast and high resolution) but averaging over multiple shots to get adequate SNR for a good DIAL or cloud property retrieval. An average value (and variance if available) is still a very scientifically useful value.

From the original paper:
It is well known that DIAL estimates of absolute humidity tend to be nonphysical within cloud structures. The values produced in clouds are typically highly variable, ranging from negative to unrealistically high values. While no comprehensive explanation of this phenomenon has been robustly demonstrated, some of our research on this subject suggests the errors in clouds are combinations of nonlinear observations (both hardware driven due to detector behavior and the nonlinearity of the observation to the water vapor), integrating over heterogeneous backscatter structure (in range and time) and ignoring the effects of the laser pulse convolution (e.g. see Hayman et al. (2023)). The reader should note that, while such errors are obvious in DIAL, this very likely indicates that all lidar retrievals in clouds will have significant unaccounted for errors.

Response:

Thank you for the feedback. The intent of the Masking section (the section containing the text referenced above) is to provide transparency about how data quality performed in the analysis presented in this paper. We agree that investigating sampling strategy is an interesting and potentially important research topic, but would likely need its own dedicated publication. We don't feel we can accurately comment on this without significantly deviating from the theme of this work. In an effort to clarify the purpose of the section and provide a better context for the masking typically conducted in DIAL analysis, we have substantially modified the opening of the Masking section. This includes more discussion about the possibility of clouds where data may be valid and a reference to Groß et al 2014. This new text is printed below for your convenience.

It is common practice to mask cloud structure in DIAL estimates due to the frequent occurrence of biases in retrieved water vapor concentration. These biases are often easy to identify because the retrieved quantities are non-physical (e.g. negative water vapor or unrealistically large values) in the cloud structures. There is relatively little published research on the root cause of data quality issues for DIAL in cloud, and the dominant sources of issues may vary between hardware architectures. Differences in online and offline sample volumes, laser pulse length convolution, detector nonlinearity, and Rayleigh-Doppler effect are common explanations for the error in most clouds; however, our own efforts to demonstrate and isolate the impact of these and other errors have been difficult to comprehensively validate experimentally. It seems entirely possible that there are additional contributing factors. It is also likely that there are some conditions where DIAL retrievals are valid in certain types of clouds and under certain conditions. For example, \cite{Gros2014} showed DIAL retrievals in high altitude cirrus that agreed well with an in situ sensor on a second aircraft over a 30 minute period. However, the analysis presented in that work was fairly limited, making it difficult to generalize even to other cirrus cloud cases. For now, it is not entirely clear what all the interacting conditions are that influence DIAL retrieval accuracy or if it is possible to know if those conditions are satisfied from the lidar data alone (meaning no external validation source would be required). For this reason it is often generalized that DIAL data is valid in clear air and most aerosol loaded conditions, but not in clouds.

Our masking routine developed here is based on the hypothesis that much of the bias seen in DIAL observations of clouds originates from averaging over heterogeneous backscatter structure \citep{HaymanAMS2023}\footnote{the presentation is available at minute 32 here https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Session/63551}. However, based on ultra-high-resolution observations of clouds by MPD (similar to that reported in \cite{HaymanSciRep2024} captured at 8 kHz and 5 ns resolution and processed to 50 Hz and 75 cm resolution) we do not believe that this represents a \emph{fully} comprehensive description of why most clouds produce non-physical water vapor retrievals. Nevertheless, masking data that produce errors is a necessary component of DIAL processing. The masking approach described here represents a heuristic approach based on the described reasoning. Although masking would likely be further improved if more comprehensive understanding of

these errors were robustly demonstrated, the approach outlined here represents an improvement over the methods we had described previously.