Reply to Reviewer 2:

Reviewer comments are in Black while our response is in Blue.

1. Introduction: please cite Wehr et al. (2023) and/or Eisinger et al. (2024) to put the products described in this paper in the full context of the mission and the other products.

Noted. This will be done.

2. Introduction: the coverage of the relevant previous literature is inadequate. Please cite papers on the existing HSRL lidars in space (Aeolus and ACDL), e.g. Flament et al. (AMT 2021), Ehlers et al. (AMT 2022) and Liu et al. (AMT 2024), discussing briefly how this work differs from your own. Regarding the algorithm, you could also cite Mason et al. (AMT 2023 "A unified synergistic...") which builds on Delanoe & Hogan (2008) and also uses an optimal estimation approach to invert the lidar signal accounting for multiple scattering.

Noted. This will be done.

3. L211: my understanding is that chi-squared is not the same as a cost function, since it normally only captures the deviations of the model to the observations (i.e. only the first term in your Eq. 9). If you must use chi-squared then the chi in L211 should definitely be squared.

Noted: Chi^2 will be replaced by J

4. Eq. 9: The use of two different types of "x" (one linear and one logarithmic) in the same equation is inconsistent with previous literature. You should define your state vector (boldface "x") as whatever is returned by your minimization algorithm. In your case (and also in Delanoe & Hogan 2008) it happens to contain the logarithm of physical quantities. So there is no need for a separate x^l vector - just define x appropriately. The forward model can still be written as F(x) as it is obvious that the first step of F will be to convert the logarithmic input values back to linear. The a-priori vector (x_a) also then does not need the "l" superscript.

Noted: The suggestion will be adopted.

Eq. 15: this doesn't make sense to me and I can't find it in Kliewer et al. An obvious problem is that the errror variance (sigma^2) of a linear variable has the units of the square of the units of that variable. But it is added to "1", which has no units! If sigma^2_xlin is the error variance of a linear quantity xlin, then the error variance of log10(xlin) is sigma^2_xlin / (xlin^2 * ln(10)^2) by the rules of error propagation, so why not use this? Since the a-priori error used in OE algorithms is often just chosen without rigorous derivation, an alternative is to specify that the error variance of the a-priori estimates of the state variables is a constant for each type of variable, indicating a constant fractional error. For example, a "factor of 2" error in xlin implies that x=log10(xlin) has an error standard deviation of log10(2)=0.3 and hence a constant error variance of 0.3^2=0.09.

There was a typo in Eq.15. The correct equation should be

S = log10(1.0+(sigma_x/mean_x)^2)

6. Eq. 16: this also seems incorrect, but for a different reason. As far as I can tell your observations and forward-modelled observations in the cost function are kept in linear space, so surely the observational error covariance matrix should be kept in linear space? It is not relevant here whether the state vector is linear or logarithmic with respect to underlying physical quantities. Also, the definition in the text immediately after Eq. 16 refers to the sigma of x, not y.

Noted: The reviewer is correct. Due to some mangled editing there are some typos in the description and equations this may have caused some confusion (see also our response to Reviewer 1). Just to be clear, our state-vector is logarithm BUT the observables (y and F(x)) are linear (along with their uncertainty). Using the log form for the observables, we feel could indeed be problematic.

7. L500: it would be useful to state the difference in speed between your algorithm and the quasi-small-angle algorithm of Hogan (2008).

Noted. A small discussion will be added. In a nutshell, the PPT (Platt+Tails) approach has a lower baseline cost but is O(N2) not O(N). So, for low to medium resolutions, the PPT approach is somewhat faster, but as the resolution increases, then the QSA approach is faster.

8. L515: There are actually two causes of tails, but this section only describes the small-angle enhancement of the Rayleigh scattering below cloud. The other occurs only in optically thick clouds and is described in numerous places, e.g. the introduction of Hogan (2008). Since the reader might be more familiar with the other mechanism, please stress here which mechanism you are talking about.

Noted: The difference between the tails treated here and “temporal pulse stretching” will be discussed.

9. Fig. B1: is this just a forward model or is a retrieval also involved? It would be better/simpler if this plot simply showed the forward model for a known profile.

There is no retrieval involved here. We will try to make this clearer in the caption and associated text in the manuscript body.

10. Sections C1.x: are these really needed since you are simply taking the derivative of some expressions? Just say you implemented analytically-derived derivatives.

Since this paper is intended to be ATBD-like and serve as an algorithm baseline reference, we think it is appropriate that they are included.

Minor comments:

We thank the reviewer for these useful editorial comments and will take appropriate action.