Reply to Reviewer 1:

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

Page 2 Line 27: To call all particulate scattering Mie scattering is rather unfortune as also stated a bit further down in the text. It is recognized to reworking the paper to use particulate consistently is a lengthy and tedious task but non-the-less strongly recommended!!!

This (unfortunate) informal convention is well entrenched within the Aeolus/ATLID ESA lidar community and used within all the other relevant papers within the special issue. We will add a more prominent `disclamer' to the text but keep using 'Mie' and `Particulate' interchangeable.

Page 5 Line 122: ATB is not defined before in the text. One could guess that it would probably mean Attenuated Backscatter. But the definition should be made at the first occurrence of the acronym.

This will be addressed.

Page 5 Line 125: "Rather, the highest index within…". This refers to the feature mask definition, but to help the reader, the meanings of the numerical values should be given here it should be noted that the feature mask values are ordered in ascending order with potential scattering strength.

This suggestion will be adopted.

Page 6 Line 134: "… are set to > 8 …" Please translate this to verbal feature type!

This suggestion will be adopted.

Page 6 Line 141: "The scattering ratio calculations…" This seems to be only a half-sentence with unclear meaning.

Noted

This text should read…"….is determined using the scattering ratio…"

Page 6 Lines 150-151: Again, feature mask entries should be given, at least additionally, in plane words.

This suggestion will be adopted.

Page 7 Line 165-166: "provides also" or "enables", one should be sufficient.

This typo will be fixed.

Page 7 Line185ff: The term "layer" here lets one think of straight horizontal layers, but in reality, the aerosol structure is often not stratiform but modulated for example by gravity waves with amplitudes of +-500 m or more. And even without gravity waves the layers are typically more wedge shaped, instead of being horizontal bands. Maybe the algorithm can cope with such cases, but this is not clear from the text. Please add more information on this.

The coarse and fine layering determination is done on the 1-km horizontal scale (although the signals used at this stage have been smoothed to lower horizontal resolutions (a few km)). So, in general, structures present at finer resolutions will be blurred out.

Appropriate text will be worked into the manuscript.

Page 8 Line 198: Step 16 seems to be missing. Please add a note that this is on purpose – if it is…

Noted. Step 16 is indeed missing from the description and Fig.2. Note, since the initial submission changes have been made in the algorithm which will necessitate an update of this section and Fig.2 in any case (a reclassification is now performed following the MS correction procedure (step 17)).

Page 9 Line 227: Using the log-form a quantity that naturally can come close to zero, like the extinction coefficient, can be problematic numerically. This introduces a pole at 0 which easily misguides the optimisation routine. Further it skews the nearly Gaussian input error statistics heavily making the minimisation of Eq. (9) not an optimal estimate any more. And the skewed error statistics leads to noise induced biases. If the whole reason for the log-form is to keep positiveness than there are methods to do constrained optimal estimates which are more robust. It is clear to the reviewer that the algorithm is, as it is at the moment and will not be changed in short term. But a few sentences more about the choice and possible alternatives should be added.

Due to some mangled editing, there are some typos in the description and equations this may have caused some confusion (see also our response 6 to Reviewer 2). 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.

As for the use of log-forms in the state vector. The reviewer claims that `` it skews the nearly Gaussian input error statistics heavily making the minimisation of Eq. (9) not an optimal estimate any more''. This is true if the a priori distribution of the state vector element is Gaussian. IF the variable is better described by a log-normal distribution, then the use of the log form (with the appropriate form of the uncertainties) is appropriate (see Q. J. R. Meteorol. Soc. 142: 274–286, January 2016 A DOI:10.1002/qj.2651). Here, the primary use of the log form is because e.g. the lidar-ratio distribution is thought to be better described by a log-normal rather than normal form (see e.g. Fig 6 of Int. J. Environ. Res. Public Health 2016, 13, 508; doi:10.3390/ijerph13050508).

In our experience, the use of the log-form for the state-vector does not present any particular challenges for the minimization. In fact, it has been shown to have benefits in the quality of the retrievals (see e.g. Maahn, M., D. D. Turner, U. Löhnert, D. J. Posselt, K. Ebell, G. G. Mace, and J. M. Comstock, 2020: Optimal Estimation Retrievals and Their Uncertainties: What Every Atmospheric Scientist Should Know. Bull. Amer. Meteor. Soc., 101, E1512–E1523, https://doi.org/10.1175/BAMS-D-19-0027.1.)

Page 10 Line 242 and 246: "…the effective radii are specified a priory by type." And "… is the a priori (linear)uncertainty assigned…" To what value? And why this value (Ref)? Maybe a table which collects all the a priory values would be appropriate.

The priors used in this work are somewhat tuned to the simulated data sets being used for the testing and may be different from those that will be used operationaly. Such a Table will be published along with the first "real" results though ! It will be pointed out, however, that for interested parties that the a priori info (along with all the algorithm settings) are accessible via the algorithm configuration files contained in the supplementary data package.

Page 10 Line 244: "It is assumed that the a prior errors are uncorrelated…" While for the observational errors it makes perfect sense to assume that different vertical bins are uncorrelated (at least for the raw signals without some averaging), this is not obvious at all for the a priory values. The atmosphere is not "white noise" and the possible deviations of a first guess from the true value no less. It is clear that it is hard to come up with sensible values just starting from scratch, and it may have serious impacts on the algorithm used for the optimisation, but this point should be discussed in more detail and a route to future improvements should be outlined.

We agree that the assumption that the priors are not correlated is likely inaccurate and merely an expedient. To improve upon this, one may envision a "boot starting" process using real observations.

Page 13 Line 345: Above or between???

Will be fixed.

Page 20/21/31: The colour scales are rather unfortunate. It is practically not possible to distinguish vales in S between 0 and 50. The one chosen for Fig. 9 is much better. It would be good anyway to harmonize the colour scales of the different 2-d charts.

Noted: The colour scales will be adjusted.

Page 34 Line 439: use? Or find?

Will be fixed.

Page 36 Line 490: "Fig ?" Reference is missing.

Will be fixed.

Page 37 Figure A1 caption: "… log-derivative approach (Grey-solid-line) and …" this should be Black-solid-line, shouldn't it? It would help the reader greatly if line legends could be added to the figures, like e.g. in Fig B5-B9!

Black-solid line is correct. Captions will be added.

Page 38 Figure B1: Here also in-plot legends would be beneficial! And the plotting style very much reminds of the pen-plotters used until the early 80s of the last centuries. The authors are strongly advised to use a homogeneous and up-to-date style for their plots!

Captions will be added. The reviewer's implications that these plots look rather "old school" is taken as a compliment.

Page 40 Line 546: "… the divergence of the forward scattered light will also be Gaussian with a divergence…" This is a simplifying assumption since the foreword scattering peak is not exactly a Gaussian! Typically, this assumption is good enough, but the phrasing here suggest that this is a mathematical truth, which is not the case.

Noted: Appropriate texts will be added.

Page 41 Eq. B8: "H(Theta_sc(z) > 0)" It is obvious what the authors try to express here, but formally this makes no sense. H is defined for the real number, but Theta_sc(z) > 0 is a comparison with a Boolean result. Even if one assigns 0 for false and 1 for true this does not work as for the common definition of H, H(0) = 1 and H(x) = 0 only for strictly negative x.

Noted: This section will we rewritten to remove the use of H.

Page 43: Again, horrible figure style. See comment above.

See our earlier response (3 comments above).

Page 45 Line 628: "calculate" -> calculated

Noted. Will be fixed.

Page 45 Line 649: Here and from the preceding text one could get the impression that Platts eta is some sort of system constant which can be "calibrated" using some higher accuracy MS-algorithms like e.g. Monte-Carlo simulations (as done for B4). But according to Platts papers neta also depends on penetration depth and optical thickness of the cloud and the variation are much larger then +- 10%. Maybe the introduction of the additional tail-function f_e takes over some of the effects of eta and stabilises its value. This should be discussed in more detail.

This is a useful observation. As it is often employed in a practical sense (e.g. CALIPSO) Platt's eta is treated as constant, but if one consults Platt's papers then one does see he considered it as a function of e.g. OT. The work presented here indeed shows that eta can be treated as a constant since, indeed, f_e, gives the system the necessary degrees of freedom. In a sense, eta_platt(z) is (approximately) equivalent to (eta_constant,f_e(z)). One could mathematically develop this idea. This will be considered for the revision.

Appendix: The reviewer had no time check that the equations for the partial derivatives of the foreword model are correct. He only hopes that they have been checked by a second person or a computer algebra program.

The derivates (as coded) were check against finite differences. A co-author has checked the derivatives as presented in the paper.