Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., author comment AC3
https://doi.org/10.5194/amt-2021-212-AC3, 2021

Reply on RC3

Frithjof Ehlers et al.

Author comment on "Optimization of Aeolus' aerosol optical properties by maximum-likelihood estimation" by Frithjof Ehlers et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-212-AC3, 2021

General comment:

First thing to note is that we were very glad about the thorough feedback of all Referees, who evidently made the effort to go through the details presented in our manuscript. This helped a lot to refine the text and we hope that we could make our statements more concise now. Especially the method of calculating the uncertainties in the MLE method was addressed by more than one of the Referees and required some changes that we want to present to all of Referees.

The procedure outlined in the last paragraph of Appendix C was relying on the simple equation alpha = gamma * beta with extinction coefficient alpha, lidar ratio gamma and backscatter coefficient beta. Motivated by comparing the obtained variances via (C2) with the variances one would get from a standard error propagation with constant lidar ratio sigma_a^2 = gamma^2 * sigma_b^2, we argued e.g. that the uncertainty in alpha cannot exceed 200 sr times the uncertainty of beta in a single bin, whatever the precise value of gamma is in reality, due to the upper bound in lidar ratio.

Now, after the feedback and some graphical considerations, the problem with this statement is, that it leaves aside any uncertainty or spread of the lidar ratio estimate itself, as it can lie somewhere in between 2 sr and 200 sr and hence should have non-vanishing variance sigma_g^2, causing another term to arise from the view of standard error propagation, increasing the uncertainty of alpha proportional to beta itself. This is why we noticed that the presented approximation was not correct, although we observed uncertainty values that matched approximately the SCA mid bin output for the SAL case for backscatter and extinction coefficients, which seemed realistic at first glance.

Following this, we evaluated the errors from equation (C2) directly against the uncertainties calculated by the SCA standard error propagation as in Flamant et al., 2020, and found good agreement, when the same Poisson noise variances are used in the S_y in (C2), see attached figure.

That means, that with (C2) we are only able to provide uncertainties for the unconstrained problem (as SCA solves it), but especially the so-obtained extinction uncertainties are overestimated and not representative of the real uncertainties, as one can see in the comparison of SCA and MLE in the simulation cases.

In lack of a representative alternative for uncertainty estimation we decided to delete the lidar ratio and extinction uncertainties of the MLE solutions in Figures 5 and 6, for now, stating that there is no representative error estimate available yet. The challenge of reporting representative uncertainty estimates is now being stressed in the abstract and conclusion and will require future investigation.

To Referee 3: It would be helpful to get more detailed feedback on the proposed "Monte-Carlo" procedure for observational data, because we are curious but did not fully understand the remark you made. However, the error estimates in section 4.3 and 4.4 are not essential for the main purpose of the investigation, because the increased precision of the MLE (and therefore its main advantage) is readily demonstrated in sections 4.1-4.4 against simulation cases and reference data. Due to the virtually identical forward model, the precision of the MLE is always higher than or on the order of the SCA.

Regarding the comments of Referee 3:

While the remarks on the uncertainty estimation have been very helpful, as described above, we also want to address your second concern on the main cause of the improvements of the MLE. The forward models for SCA and MLE are virtually identical, besides the included variable transforms. That implies that the only qualitative difference between the methods is the constraints. These are now acknowledged as a-priori knowledge, making a stricter differentiation between an explicit cost function term and a-rpiori knowledge in general. Without these constraints, the MLE would lack the physical information to identify signal noise (see equations A7 and A8 in Appendix A) in the first place. In order to illustrate this influence more convincingly, we attached a comparison of real data case 1 processed with and without constraints. This underlines, that the dust plume (SAL) retrieved with MLE without constraints is disrupted similarly to SCA results in all aerosol optical properties. We hope to have made suitable modifications to to clarify this better, please find them outlined in the answers to your comments below. All noted grammar and spelling errors have been adopted without explicit mentioning.

P3 L77. A positivity constraint on quantities where zero is a valid and common expected value will produce a bias. The authors mention this in the context of the SCA algorithm, but this is also a concern in the new algorithm. It might be better not to include the positivity constraint on backscatter and extinction in the new algorithm. Have the authors tried their MLE method without the positivity constraints?

Yes, we tried this. The MLE method without any constraints shows no advantage over the SCA, since the observed signal values can be perfectly fitted to. Without the zero-flooring, the SCA produces oscillatory extinction results as in Figure 10 in https://earth.esa.int/eogateway/documents/20142/37627/Aeolus-Data-Innovation-Science-Cluster-DISC-Level-2A-user-guide.pdf. Comparably bad solutions are obtained using MLE without constraints.

We added after equation (14):

"It is important to mention that it is only the a-priori knowledge in form of the box-constraints that makes the MLE solution distinct from the algebraic SCA solution (without zero flooring, see section 6.2.2.1 in Flamant et al. 2020), because this algebraic solution corresponds to the exact signal values in y and therefore to J_{obs}=0, which is the global minimum to the unconstrained counterpart of problem (14)."

p10 Box constraints. I think its possible that adding constraints has a smaller impact on improving the results than the other improvements: using raw measurements and uncertaintites in a coupled retrieval.

The forward model equations used by SCA and by MLE are virtually identical and differ only in terms of some variable transforms between optical depth, lidar ratio, backscatter and extinction, see section 3. So it is indeed the a-priori knowledge in form of the constraints that causes the improvements.

Additionally, the constraints have potential negative consequencs: (1) that they could lead to bias (similar to what happens in SCA) and (2) that they make propagation of the uncertainty very difficult. Have the authors attempted the MLE retrieval without adding the box constraints? It would be useful to compare results with and without using the box constraints. If they are just as good, eliminating the box constraints would eliminate the two issues mentioned above. If they are not as good, the comparison would give a clearer view of the impact of the different improvements.

This is a good point and has not been made entirely clear in the manuscript either, but as mentioned above: Solving the MLE without box-constraints is essentially a more complicated way to converge to the algebraic SCA solution (without zero-flooring), which is not favourable.

P19, L469 "This is mainly due". Similar to the previous comment. This statement is made without any analysis to suggest how the different algorithm features dominate the improvements. The authors should assess the impact of the box constraints separately, to support this statement that it is the dominant cause of improved results.

For the above mentioned reasons, the SCA essentially is this benchmark. So, we do not intend to include an analysis of the "MLE without constraints", as it would just illustrate the oscillatory behaviour outlined in section 6.2.2.1 in Flamant et al. 2017.

P6 Eqs 2 and 3. P21, Eqs A4 and A5. Very little is said about the cross-talk constants C1, C2, C3 and C4. The top of page 7 seems to imply the constants are known, but in fact they are very challenging to determine and have significant uncertainty associated with them. More information is needed about how these are determined and what are typical values for Aeolus. This is also important for the discussion of the correlated errors and for understanding the magnitude of that problem.

We agree, the typical magnitudes for C1...C4 should be added. For the calculation of the calibration coefficients please see the L2A ATBD (Flamant et al. 2020, updated 8th Juli) section 6.5 and the stated reference document RD25 "Generation of AUX_CAL Detailed Processing Model Input/Output data definition" available at https://earth.esa.int/eogateway/documents/20142/37627/Aeolus-AUX-CAL-IODD-DPM.pdf . To add this essential information in the text, we included just above section 3.1:

"The instrument is calibrated with measurements from dedicated instrument calibration modes (Reitebuch et al., 2018a) and the cross-talk coefficients C_{1...4} are determined according to Flamant et al.,2020 and the procedure in Dabas, 2017. At (p,t,f) = (1000 hPa, 300 K, 0 MHz) C_1 and C_4 equal 1 by definition. The other coefficients then typically take values about $C_2 \approx 0.5$ and $C_3 \approx 1.25$. For the rest of this work, we assume the calibration as known and do not include the contribution of calibration errors in the results. The calibration cannot be perfect in reality, but is likewise input to all algorithms, which guarantees a fair comparison of retrieval precision in Sections 4.3 and 4.4."

We also include in the abstract now:

"The increased precision of MLE with respect to SCA is demonstrated by increased horizontal homogeneity and better agreement with the ground truth on real data cases, though proper uncertainty estimation of MLE results is challenged by the constraints and the accuracy of MLE and SCA retrievals can depend on calibration errors, which have not been considered."

p10-11 L294-304, two approaches to error quantification. While in general I like the idea of analytical propagation of errors using sensitivity analysis, the propagation is seriously challenged by the difficulty of representing the impact of the constraint in the propagation equation. I'm not very convinced by this method of rescaling the variance and assuming the correlation is unchanged (Appendix C, near P23 L575). Can it be explained more clearly why the correlation matrix should remain unchanged? Also, using the lidar ratio constraint to (potentially) reduce both the extinction and backscatter error bounds separately gives an impression of circularity. The authors should add rigor and validation to make this method more convincing.

see statement of the author's

P25 Figure D1. Although the Monte Carlo methodology is discussed, it seems that the results of the Monte Carlo approach are only shown in Appendix D without any analysis or discussion. These should be promoted to main text and properly analyzed.

The uncertainties in the Figures 2 and 3 are from Monte-Carlo assessment, which is in the main text already. We now refer to it explicitly with having changed line 295 to say "A Monte-Carlo approach as in Xiao et al. 2020 is applied to classify the uncertainties in simulation results in sections 4.1 and 4.2." We also promoted modified versions of D1 and D2 to the main text in conjunction with a more detailed discussion, see coming answers below.

Both methods are available for the simulation cases, so I'd like to see a comparison of the analytical (Appendix C) method against the Monte Carlo results for the simulation cases, which might bolster my confidence in the analytic method.

see statement of the author's

It would also be useful to compare the propagated uncertainties with and without the adjustment for the constraints, to see how big this impact is.

Typically, the extinction uncertainty estimate is reduced because the uncertainty of the extinction is usually much higher than 200sr times the uncertainty of the backscatter.

The authors say a Monte Carlo approach is not feasible for the measurement cases, and I think this refers to the fact that they cannot vary the true measurement error. However, it is certainly possible to simulate measurement noise from the measurement error covariance matrix, S_y. Using these simulated measurement errors in a Monte Carlo approach would give an accurate view of how S_y is propagated through the retrieval, more immediately convincing than the method in Appendix C. So, going further, I'd like to see a Monte Carlo propagation of S_y, which could be used to validate or replace the current propagation.

We cannot follow this point well. What we have is the Forward model y = F(x) going from the state space x to measurement space y. Now, what you suggest is to propagate y errors through y = F^(-1)(x), which contains the implicit function F^(-1). Finding this implicit function locally is the problem we solve time-consumingly via optimization already. So, if we understand correctly, you suggest to estimate the error via multiple runs of the minimization procedure with artificially added noise from S_y to the already noisy signals S? This artificially increases our noise amplitude and makes it necessary to run the minimizer (at least 20 times) more often, which we find to be unfeasible effort for the operational product.

P17, Figure 5 (also figure 6). I can't understand which error bounds are shown in Figure 5 and Figure 6. The descriptions "lower error bound" and "upper error bound" are confusing because they don't match any description in the methodolgy, and they are possibly inaccurate as well. Another terminology should be seletected, preferably one that reflects the descriptions in the methodology section. The two sets don't seem to be the "two approaches" introduced on P10-11. Are they instead two formulations of the analytic propagation of errors using different characterizations of the measurement error S_y? Is the "upper error bound" the propagated uncertainty from Appendix C? (If so, calling it an upper error bound is particularly problematic since the Cramer-Rao inequality characterizes a lower bound, not an upper bound). And how is the "lower error bound (Poisson)" calculated? I don't see that in the manuscript.

see statement of the author's

If indeed the one labeled "upper error bound" is the one produced in this analysis, and the one labeled "lower error bound" is the standard one for SCA and midbin-SCA, then another emergent theme of this analysis is that the uncertainty of the standard algorithms is inaccurate as well. This should be highlighted in the manuscript as another primary impact of the new algorithm.

That is right. Finding the most representative errors in SCA and SCA MB estimates is under current investigation at Meteo France. That means, that the presented Poisson error estimates for SCA and SCA MB are also to be validated. This is why we decided to use signal uncertainties calculated as in Appendix B in the first place for MLE.

P1 L8 "algebraic inversion scheme to a (partly) ill-posed problem and therefore sensitive to measurement noise". The sentence should probably be reworked. "(partly) ill-posed" is not really informative, and even a well-posed inversion is sensitive to measurement noise.

This statement was based on the work of Pornsawad et al., 2008: "But From a mathematical point of view Eq. (1) is a Volterra integral equation of the first kind, which is ill-posed." We find this wording misleading as well and thus decided to drop "ill-posed" throughout our text.

P3, L78-79. I think the sentence should be split up into multiple sentences that clearly list the factors that led to poor results in the SCA (there are at least two, I think: the correlated noise and the incorrect removal of negative extinctions) and the factors that lead to improvements (coupled retrieval, using the measurement error covariance matrix in the retrieval, box constraints), and to explain their causal relationships. In particular, I believe it is incorrect to suggest that the box constraints are responsible for eliminating the correlated noise issue, as this sentence seems to say. Instead, I think the bias is avoided by using the measurement channels and their error covariance matrix in a coupled retrieval.

We keep this sentence as it is but added the respective reference to the zero-flooring and averaging operations in SCA within Appendix A. Without the box-constraints, the MLE would not 'know' which fraction of any signal value likely corresponds to noise, therefore, in (A7) and (A8) the MLE would be 'blind' and not be able to improve anything.

P4, L130. How is the grid spacing chosen? What are the typical values?

The range bin sizes are typically coarser at high altitudes and finer close to the ground but change several times per orbit and area of interest, which makes it difficult to generalize. For better readability we changed this sentence to "Individual, vertical range bin sizes can be independently varied between 250 m and 2000 m in steps of 250 m, while remaining limited to a total number of ACCD rows and hence vertical range bins of 25 per column."

p7 Eqs 9 and 10. The minus sign in the denominator does not appear in Flament et al 2017. The rest of the equation looks similar but with a few more algebraic steps. I think it is otherwise correct, but I would like the authors to double-check to be sure.

Thanks for spotting this error; the minus sign has been corrected and the equations have been double checked.

p7 Eqs. 9 and 10. It would be helpful to have a sentence or two summarizing the derivation of the quoted equations; for example, an explanation that the equation includes an expression that explictly integrates the transmission over an extended vertical bin.

For better reading flow we changed the line above these equations to say: "The following approximations for the range corrected signals (5) and (6) are made by using the mean bin properties from above, see equations (6.35)-(6.36) and definitions above in Flamant et al., 2020:" We hope that clarifies? We are sorry but we cannot repeat all derivation steps from Flamant et al. as this would not add value to the manuscript. The message here is, that the identical SCA forward model equations are used (with transformed variables).

p8 near the the end of section 3.1, it would be good to have a sentence or two discussing the anti-correlated noise that's mentioned in the Appendix. This seems like an important point that is referred to both in the introduction and later, so it should be described in the main text with enough detail so a reader knows what it is about.

It is not too important and was purposely placed in the Appendix to compress the main text of the manuscript. The main insight after the algebra in 3.1 should be that the backscatter and extinction depend on both Sray and Smie at the same time.

P9 L232. I'd like to know more about how the box constraint is implemented and specifically whether it may negatively impact results. Could it potentially bias the results (by perhaps producing a disproportionate number of solutions near the boundaries) or affect their precision of the solution (via a variable transform that changes the gradients near the solution and near the boundaries)?

We are no experts in this either, but these details of the L-BFGS-B algorithm can be found in the work of Nocedal et al. and Zhu et al. (as referenced) on http://users.iems.northwestern.edu/~nocedal/lbfgsb.html, which is underlying the scipy.optimization package.

p9 242. "future developments". This is really more of a response to another reviewer, but I just wanted to say that I support the authors' decision to perform single profile retrievals without including horizontal scene smoothing. It makes sense to explore the simpler solution first. Furthermore, not requiring scene continuity for the algorithm means that scene continuity can be used as a check on algorithm performance.

We thank you for your remark and fully agree with this position, as otherwise no conclusion could be drawn from the comparisons in Figures 5 and 6.

p9 L250 As I mentioned, I believe that using the measurements and their appropriate measurement error covariance matrix is largely responsible for avoiding the bias due to the correlated errors in the cross-talk corrected signals. Do the authors agree? If so, this paragraph might be a good place to highlight that.

On the contrary, we will need to highlight that the a-priori information in the forms of the constraints is responsible for the noise suppression capabilities. This we hope to have accounted for with the additional text added according to comment on P3 L77.

p10 L272-275. Is there anything in the current algorithm that addresses the bias from partially filled bins? If not, is the current algorithm compatible with the strategy implemented by Flament et al. 2017?

No, we do not follow the strategy of the ICA, that is why SCA is used as a benchmark. While the approach reads promising, the ICA has essentially too many degrees of freedom to yield useful results and is not currently developed any further (see Aeolus L2A user guide in ESA resources) due to its very poor performance. In the data product, the only columns that seem reasonable are basically almost unchanged copies of the SCA results. Consequently, no effort has been made to tailor this implementation to the ICA. Already with the SCA, one could criticize that the effective resolution of extinction coefficients is expected to be coarser than the range bin size itself.

p10 L290-292. There appear to be three rather weak arguments here to justify not using an a priori covariance matrix such as called for by OE. The weakest part of the argument, in my opinion, is saying that the algorithm has no prior information or regularization. In fact, the constraint is prior information and does provide regularization. If the authors believe the influence of the constraint is significant (which I think they do, because they attribute the improved results largely to implementing the constraint), then even suggesting that this is a weak constraint would not be justified. Therefore, the authors should acknowledge that the constraint plays the role of prior information in the retreival, but they chose not to cast this constraint in the terms required by optimal estimation, because that would require a different (probabalistic) form for the prior that isn't compatible with the desired form of the constraint. In my opinion, that is a reasonable reason to use constrained MLE rather than OE. (However, the choice leads to difficulties in working out the correct way to calculate the impact of the constraint on the uncertainties, which I discussed in the "Major" section).

Do you rather mean lines 231-232? We fully agree with your understanding of the matter. What we meant was that we do not have an explicit term in the cost function. We highlight the importance of the box-constraints but also added after (14): “[...] with box-constraints on lidar ratio, which is prior information that cannot be exactly represented by OEM.”

p10 L295. Some clarification would be helpful here about the simulated measurement noise. Is it simulated as gaussian with the variance determined from Appendix B, or is it simulated with various separate error components? Which components are included?

The information on which error components on the signal noise the simulation results include can be found in the beginning of section 4. These noise components are to the best of our knowledge modelled realistically, not just added up variances.

p12 L347-349. While I certainly agree that filtering negative results would cause a high bias that is worse in bins with low SNR, I can't clearly follow the explanation in this part of the text and I don't see a particularly clear indicator that they are triggered by the shift from coarser high-altitude bins to finer low-altitude bins as implied by the text. Can the explanation be clarified?

According to Reviewer 2 we promoted a modified version of FIg. D1 into the main text, which we now refer to for clarity of the statement. You can see the locally highest extinction coefficient biases at 2 and 13 km altitudes in Fig. D1 exactly where the range bin settings change. An exception is the zeroth bin of course, since it is always normalized to zero extinction in SCA.

p12 L354. The indicator from the averaging kernal is a good idea. Is there a quality flag in the data product related to this? It would be nice if this indicator were included in the plots to show which bins are not trustworthy. This would be particularly useful in Figure 4, for instance, where I am curious to know if the bins below the lofted aerosol are reliably retrieved. This curiosity is fueled by the fact that CALIPSO's 532 nm data frequently misses the aerosol at the bottom of attached layers due to attenuation, making them look lofted. The HSRL capability should act to prevent this problem given sufficient SNR, but on the other hand, attenuation at the shorter wavelength of 355 nm would be worse. So, to be sure, it would be good to see some indication from the retrieval algorithm that the lack of aerosol below the apparent plume bottom is reliable.

The values that cannot be retrieved are padded with -1 in the data product. Usually, this regards only the lowermost extinction bin. Regarding the bins below the lofted layer: We do of course see increasing error estimates with increasing distance from the satellite but we fear we cannot say if the layer does only appear to be lofted in both Aeolus and CALIPSO data or if it is lofted for real.

p12 L354. "The extinction bin closest to the ground cannot be well retrieved." Is this true in general, or specific to this case?

True in general, due to the set of the used equations. We added the word 'generally' to the text for clarity.

p12 L357 "Otherwise". I'm confused by this sentence, and not sure if I'm confused by content or just the wording. Does "otherwise" indicate the strategy of taking the mean lidar ratio? I agree that taking the ratio of the mean extinction and backscatter is a better strategy than the mean lidar ratio and will give potentially different results, since lidar ratio for small extinction and backscatter is not as reliable as when the SNR is higher. But why is the mean lidar ratio contaminated by the first guess? And how does taking the ratio of mean extinction and backscatter avoid the influence of the first guess?

This overlaps with a point that Reviewer 1 made as well. So the problem is the following: Whenever the retrieved aerosol optical depth L_p is very low, the influence of the lidar ratio estimate on the cost function becomes increasingly insignificant, until L_p becomes zero and no lidar ratio can be provided at all (these are not included in the statistics). Wherever L_p is very low, the lidar ratio has a high error estimate and some may remain close to the first guess, which then seems to act as an 'implicit a priori' (see https://doi-org.tudelft.idm.oclc.org/10.1029/JD095iD05p05587, end of section 7). What our paragraph wants to say is that, if lidar ratios were averaged without weights, then the first guess would contaminate the statistics. In the shown simulation case, this would only affect the statistics when only bins with retrieved backscatter coefficients below 5*10^(-2) (Mm sr)^(-1) are considered.

So eventually we propose to replace the sentence "Otherwise, the first guess of $\gamma_{||,p}=60$ sr would contaminate the statistics for MLE, e.g., $\text{mean}(\alpha_p/\beta_{||,p})$ would be biased towards the first guess." with "[...], in order to disregard the influence of bins with nearly vanishing aerosol optical depth, for which no reliable lidar ratio can be retrieved."

Additionally, lines 360-361 and lines 380-381 are omitted, because they are more confusing than helpful.

Figure 2. What is the explanation for the low bias in the median lidar ratio from the MLE for the lowest bins?

The pdf of MLE backscatter is asymmetric here, due to the positivity constraint. Hence, the median is not equal to the mean.

Figure 2-4 and D1: The line plots are so small that I can't see important information. The data are particularly important below about 2.5 km where there is significant aerosol, but this is a very small portion of the plot and not readable due to its size and the closely spaced grid lines. Specifically, the lowest bin is called out in the text, but it is so small I cannot see the data or error bar for that lowest bin in the line plots. Please include inset boxes to show the line plot data in the lowest 2.5 km, or a second set of line plot figures that show only the lowest 2.5 km, or in some other way improve the visualization of the lowest 2.5 km.

This is a really good point for consideration. So we now adjusted the font size in the mentioned figures to match better with the text and also provide and discuss an overhaul of figure D1 in section 4.1, in which inset boxes make this information on biases and errors accessible.

p14 L362. The reduced uncertainty from the new retrieval compared to the SCA should be discussed and quantified. I think the new retrieval probably produces usefully smaller uncertainty, but the fact that the figures are so tight makes it nearly impossible to see the region where there is significant aerosol below ~2.5 km, and I can't even tell if the error bars are smaller than 100%. A discussion of uncertainty results is just as important as the mean tendency of the profile, because, for example, a profile that "looks right" but is indistiguishable from 0 everywhere due to its uncertainty would be rather useless.

According to the remark above we considered your feedback and added a more detailed discussion in 3.1 by including: "In order to better illustrate this, the bias $(\text{mean}(x^{*})-x_\text{true})/x_\text{true}$ and relative error $\text{std}(x^{*})/x_\text{true}$ for all retrievals with respect to the true profile are presented in Fig. D1. Here, the maximum backscatter coefficient bias in the aerosol layer below 2 km is reduced from about 43\% with SCA and SCA MB to 27\% with MLE. This bias seems to be triggered by the refined range bin setting below 2 km. For the origin of this bias we can think of two causes: Firstly, the violation of the hypothesis of uniformly filled bins due to the strong gradient in aerosol concentration and, secondly, the non-linear way the backscatter coefficient is calculated from $\beta_{||,p,i}=Y_i\beta_{m,i}/X_i$, because here $mean(\beta_{||,p,i})$ will become biased high increasingly with increasing uncertainty of X_i. The relative error in backscatter coefficients is consistently lower for MLE compared to SCA; In the most interesting area below 2 km the relative error in backscatter coefficients reduces to 50\%-30\% with MLE compared to 120\%-50\% for SCA, while MLE performs only slightly better than SCA MB."

and

"The MLE retrieves the least biased extinction coefficients over the whole profile with standard errors up to a magnitude smaller than SCA midbin product, see Fig. D1. Retrieved extinction coefficients are all biased high in the area below 2 km, with maximum bias of 500\% for SCA, 110\% for SCA MB and 70\% for MLE. Between 1.5 and 0.5 km altitude, the biases are comparable in magnitude with about 30\%. Concerning the relative extinction errors, an improvement by a factor of about 1.5 to 2 in comparison to SCA and SCA MB is achieved by MLE in the lowermost 2 km. Though the relative error is on the order of 100\% or greater in all cases."

*P14, L366. Is this number a typo? exp(-2*0.4) = 0.45 not 0.38.*

This has to do with Aeolus viewing angle of about 35 degree off nadir. This means exp(-2*0.4/cos(35deg)) = 0.38 from the instruments point of view.

P14, L369-379. It seems that it's not just delayed (i.e. slow decay below the cloud) but the cloud is smoothed into the regions both below and above. In other words, it looks like the effective resolution of the extinction is much less than the backscatter, which make sense since it takes at least two measurements to calculate a derivative. Is there any attempt to calculate backscatter on the same coarser resolution as extinction to produce the lidar ratio?

The phrasing of delay regards the curtain plot in row 3 of Fig. 3 in which you can see that the onset of the cloud is positioned at the ground truth for MLE but is delayed by a bin in SCA. We now properly indicate this in the text. The slow decay below the cloud must be a result of the zero-flooring and the noise in SCA, as we cannot see it in SCA MB and MLE. MLE relies on the same forward model and should therefore have identical effective resolution, but due to the limitations of MLE as compared to OEM, we have not determined any effective resolution.

P15, Figure 3. It appears that the bias in the mean backscatter below the cloud is actually worse in the MLE than in the SCA and SCA-midbin results. This should be discussed.

By promoting figure D2 into the main text, we also added some more detailed discussion and included this observation. "A distinct feature of the MLE result is the bias of up to 100 \% between 2 and 8 km altitude, which might be introduced by the positivity constraint in conjunction with high noise. However, the bias in the area of interest below 2 km is found to be the lowest in the MLE case."

P19, L449, what causes missing values in SCA midbin?

These values are flagged out in the operational processing since they are negative. This is the case only in this figure, but has been suppressed in the generation of the end-to-end simulation statistics. For future reprocessing campaigns it is planned not to flag negative values anymore.

P19, L458, Lidar ratios aren't shown in the figure for the Real Data Case II, as in other cases. Better to show them, if possible.

The individual lidar ratio estimates on single bins are basically so noisy in the SCA cases, that this would be of no value for the reader (with zeros and negative values also due to this backscatter-extinction delay problem), therefore we provide only the calculation. We referred to this noise problem in the text already.

P19, L459-460, Is this comparison between copolarized lidar ratio from AEOLUS to total (co- and cross-polarized) lidar ratio from Polly? This is not the best option. Since Polly is also sensitive to polarization, wouldn't it be possible to calculate copolarized lidar ratio from Polly for a more direct comparison? In any case, it should be clearly stated what's being compared.

The Polly results provided by Holger Baars do mimic the Aeolus range bin settings and the co-polarized lidar ratio that Aeolus sees. We replaced all mentionings of lidar ratio in the text with co-polarized lidar ratio, because this was not made clear before.

P19, L461. It seems strange that the quoted uncertainties are from the Poisson method that the authors are hoping to replace. It would be better to quote uncertainties from the method that the authors think are more represenatative.

see statement of the author’s

P 20, Figure 6. Is there an uncertainty bound available for the "ground truth" (which is *also a retrieval)?*

We are sorry, but this uncertainty has not been provided.

P20, Eqs A2 and A3 are difficult to mentally derive from the previous step. It would be helpful if more intermediate steps were added to make it clearer how the equation is derived. The appendix is a good place to do this.

This line has been edited the same way as in comment on p7 Eqs. 9 and 10.

P22, L538. This point about anticorrelated error is interesting and informative. However, some parts of the discussion are confusing. For instance, Eqs A7 and A8 show that an error spike in one measured channel gets distributed in an anticorrelated way into the corrected signals. But how does it follow that there is correlation (or anticorrelation) in the errors in backscatter and extintion? It seems logical that if the errors in the two corrected channels are anticorrelated, the ratio would tend to be biased low, so the backscatter would tend to be biased low. However, the error in the extinction would not be correlated with it, because extinction derives from gradients in just one of these corrected channels.

What we meant to express here is rather that:

"This means, e.g., that one noisy value in the Mie (or Rayleigh) channel disturbs both backscatter and extinction coefficients. That also implies that if an unphysical, negative value is obtained in one bin for backscatter (or extinction) in the final result, then the value for extinction (or backscatter) is definitely disturbed as well, whether it lies within physical bounds or not."

This apparent additional information, which lies below the surface so to say, is not used after the SCA has been applied once. The text above now replaces: "Thus, whenever a negative extinction coefficient is found in bin i, this indicates a spurious estimate of the backscatter coefficient as well (although it might appear to be well in physical limits) and vice versa."

P23, L557, I suggest replacing "in spirit" with something more informative, such as "except with an adaptation to account for the impact of the constraint"

We decided to drop this phrase.

P23, L561. It's not true that no a priori knowledge is imposed. The box constraints are a priori knowledge.

Thanks, we changed "a-priori knowledge" to "explicit a-priori term" to be more concise.

P25 Figure D1. "Bias" and "standard error" should be defined in equations.

This has been promoted into section 4.1. See answer to p14 L362.

Title. The title would be more informative if it contained the word "aerosol". For example: Optimization of Aeolus Aerosol Optical Properties Products by Maximum-Likelihood Estimation

This is a good point that was also made by Referee 1, so we aim to change the title to indicate that out investigation does not concern optical properties possessed by Aeolus.

P1 L1. Typically "embarks" is not used this way, and used only for the much narrower circumstance of a person getting on or off a vessel or starting a journey. It could be *replaced with "includes"*

We decided for “carries”.

p3 L69-70. I'm not sure I understand what is meant by "well located". Does this mean retrieved at finer resolution? (Perhaps not, because if so, it still doesn't quite make sense. I understand that extinction and backscatter are retrieved simultaneously on the same grid, but the effective resolution of extinction and lidar ratio will always be less than the finest possible backscatter resolution because it takes at least two measurements to determine a derivative.)

We modified the sentence to “Thus, the particle extinction may occur only where there is backscatter and [...]”

P4, L130. Consider adding "irregular grid" to this description. Perhaps here: "in steps of 250m, to produce an irregular grid with a total number"

We updated this mentioning individually adjustable range bin sizes.

p7 Equations 5-8. I think the variable names and subscripts could be chosen to better indicate which set of variables are the raw signals and which are the cross-talk corrected signals. It's particularly confusing that the pair with mixed Rayleigh and Mie components are subscripted "ray" and "mie" while the pair where they are actually separate is generic with no mnemonic. Using p and m subscripts for the corrected measurements in 5 and 6 might help. I would also suggest renaming the variables in 7 and 8 to avoid the ray and mie subscripts on the raw channels (although I know that suggestion might be more controversial because it's based on historical precedent with this kind of instrument).

We fear that having inconsistent variable naming conventions compared to the main reference (Algorithm Theoretical Baseline Document, Flamant et al. 2017/2020) would further complicate understanding. This way, it is mostly aligned. The only difference is the lowercase s for signals, so to circumvent confusion with the error covariance matrix S_y (capital).

p8 L227, I believe the meaning will be clearer if you remove "to account for" and instead say "because additional noise contributions, such as ..., are likely to dominate over the Poisson noise"

The other contributions are not likely to dominate over Poisson noise, but rather to increase it noticeably and to smear it out in a quasi-Gaussian fashion, so we prefer to keep the statement as it is.

p9-10, L261-263, I had trouble understanding this sentence until I read the Appendix. It might be clearer to move "scaled" and break the sentence into two: "Here we use the variance measured at 2.9 km resolution, scaled to approximate the noise level in the 87 km bins. This approximation assumes the scene is homogeneous so that all the variability is due to measurement error."

Shorter sentences are always favourable, so we adapted the sentence.

p10 L272-275. I found these few sentences very confusing. I think the authors are saying the extinction bias due to underfilled bins tends to decrease the measured range of lidar ratio, but that this is not a compelling reason to reduce the lidar ratio upper bound in the algorithm, so they end up ignoring the bias found by Flament et al. 2017. If the effect is ignored, then these sentences are somewhat of a distraction and could simply be deleted. If the authors feel that it is important to keep these sentences about the bias due to *underfilled bins, (and if my interpretation is correct), I think the readability could be improved by (1) signalling the contrasting thought using "On the other hand" (or "In contrast") instead of "Additionally", (2) using "aerosol partially filling a range bin" instead of "different hypothesis on the distribution of aerosol layers within a range bin" (3) specifying "underestimate" instead of "alter" and (4) correcting "co-polarized particle backscatter coefficients" to "co-polarized measurements". That is, the bias due to underfilled bins is a bias to extinction, not to particle backscatter coefficients.*

We changed the text to: "On the other hand, as shown in Flamant et al. 2017, aerosol partially filling a range bin can easily underestimate the true particle extinction retrieval results by a factor of 16. The applied bounds need to account for these forward model errors by an extra margin."

p10 L288. Forty or forty thousand? Forty seems more likely, in which case three digits after the decimal point for an integer is a strange instance of false precision (should be just "40"). Or if forty thousand, it should be a comma not a point "40,000" - but in that case, 40,000 is a crazily large number of iterations. What is the typical number of iterations actually required for convergence?

When just one profile is considered, as in problem (14), then the number of iterations is on the order of 50 (roughly 0.5 seconds on an office laptop) to achieve convergence (an average cost function value per bin below 1). In practice, however, we are interested in solving problem (15) for all profiles simultaneously. In this case about 5000 iterations are required (roughly 45 seconds on an office laptop) to let the cost function decrease in the same way for a whole orbit (roughly 450 profiles). This information has been added to the manuscript.

So essentially, solving (15) for the whole orbit saves us time. Though, on second thought there must be a better converging formulation, but we have had no reference on this. We made an attempt to reformulate our model in terms of log(L_p), but the convergence speed did not increase noticeably. It is very likely that a more advantageous scaling exists since the L-BFGS-B algorithm is not scale invariant. Unfortunately we haven't found it and are currently tuning parameters to enhance performance.

p11 L305 Consider inserting Equations C2 here. The sentence could read "This relation is inverted to produce Eq. C2", and deleting the terminology "Moore-Penrose pseudoinverse" in the main text, since it's more understandable in the Appendix where it is explained more completely.

We dropped the phrase regarding the Moore-Penrose pseudoinverse and changed it into a reference to equation C2.

p12 L339. "optically thin" rather than "thin"

Adopted.

Figures: I hope Figures 2-4 will be larger with proportionately larger text. They are quite difficult to read.

This has been taken care of.

P14, L383, I'm not sure what "locks extinction and backscatter to appear colocated" means. Does it mean something like "MLE retrieves extinction and backscatter at the same effective resolution"?

We changed it into "forces extinction and backscatter to appear colocated", i.e. to appear together in the same range bin.

Figure 4 caption. "CALIPSO Feature Mask". Better to use the official CALIPSO product name and be specific: is it the "Vertical Feature Mask (VFM)" or the "Atmospheric Volume Descriptor (AVD)"?

We are sorry for the sloppy notation, we meant the vertical feature mask (VFM) and corrected the mentionings thereof in the text.

P17 L408-409. Delete "and it might be horizontally and vertically". It is disrupted, not might be.

Done.

P18 L419. I suggest using "copolarized" rather than "co-polar", to be consistent with how you have described it earlier.

This is now corrected for.

P18 L427. Explicitly include copolarized lidar ratio in this sentence: "expected values of copolarized lidar ratio of 80sr - 120sr for desert dust". It would be good to avoid the possibility of these expected values being taken out of context and mistaken for the more usual non-polarized lidar ratio.

This is now corrected for.

P18 L431. Replace "all values" with "expected values" and replace "the estimated error ranges" with "the error ranges estimated from Eqs C2 and C3" (or some other phrase to make it clear that it's the error ranges calculated from the new methodology. They do not fall within the "poisson" error bounds.)

The MLE error estimates have been deleted from the figure and the text.

P19 L449. ":Hence, " is not the right connector since it implies causality, and consequently I'm not exactly sure I know what is meant. I think the authors are only explaining what was meant in the first part, so perhaps replace with "; that is, ".

Is corrected for.

P19 L470. Replace "with this" with "in addition". Again "with this" is usually used idiomatically to indicate a causal relation.

Here we actually mean to indicate a causal relation, because it is only by the a-priori knowledge of constraints in the first place that noise can be detected and suppressed in a second step.

P19 L472-474. The flow of this paragraph is interrupted. I suggest moving these sentences about how the coupled retrieval improves backscatter to the earlier spot at L469 immediately after the statement that backscatter is improved along with extinction. The information about the constraints and the anti-correlated noise are separate thoughts.

We agree and moved the sentence on the reasons for the improvements to the end of the paragraph.

P21 L499. "the properties above" doesn't make sense here, since no properties have been defined yet in the Appendix.

Has been changed to “the properties in section 3”.

P21 L521, I suggest replacing "optical depth/extinction" with "layer optical depth". Dropping "extinction from this sentence is no loss since the relationship with extinction is described in the very next sentence.

Was adapted.

Please also note the supplement to this comment:
https://amt.copernicus.org/preprints/amt-2021-212/amt-2021-212-AC3-supplement.zip