Response to Reviewer 2.

Thank you for your positive and constructive comments. Your time and interest are greatly appreciated. Answers to your specific comments are given below in blue.

Reviewer 2 states:

I should start by saying, that this is an excellent and deep research. Calculating the the aerosol extinction coefficient is a challenging task, particularly in the presence of noise. The authors convincingly demonstrate that the Optimal Estimation (OE) technique reduces uncertainty in the calculations compared to traditional analytical methods. Additionally, the approach enables the determination of the optimal height resolution for different altitude ranges. Another key strength is its ability to account for both random and systematic uncertainties, including calibration errors.

Overall, this is a high-quality scientific manuscript that is certainly publishable in its current form. However, while certain aspects may seem self-evident to the authors, readers less familiar with OE may benefit from additional explanations.

Technical comments

1. Eq.4. Symbol “C” is the same as in Eq.1?
No, it isn’t. Thanks for catching that. It’s been changed in Eqn 1 to a unique symbol, lower-case *k*.

2. p.7 ln.18 “If the residual is approximately one or less, then the solution agrees well…”. Would be good to explain or provide a reference.
In the revision, more explanation is added, like this: “Note that the actual error in each element, y-F(x), is balanced by the uncertainties $\mathbf{S_y}$. Therefore, if the residual term is approximately one or less, that indicates the solution agrees well with the measurements, within the measurement uncertainty.”

3. p.9 ln 19. “the prior profile of backscatter is taken to be zero”. Probably needs explanation, why it is zero. Just wonder, if a prior profile can be calculated from standard analytical approach.
More explanation is added to the revision. It now reads like this:

> *Specifically, the prior profile of lidar ratio is taken to be 50 sr with a one-sigma standard deviation of 35 sr, (i.e. 95% confidence the lidar ratio falls between -20 sr and 120 sr for a normal distribution). For aerosol backscatter, the prior is taken to be zero with a one-sigma standard deviation of 0.015 km^{-1} sr^{-1}. Aerosol backscatter can vary over many orders of magnitude. This standard deviation covers a large portion of the range of values seen in many years of airborne HSRL-2 data. Of course, lidar ratio and aerosol backscatter are not distributed normally, but this setting for prior uncertainty is large enough that the shape of the distribution is unimportant. Although there can be no negative aerosol backscatter values, choosing zero as the prior is helpful since the prior will come into play primarily when the measurement signal is insufficient to constrain the results; that is, when aerosol bacskcatter is near zero. Since the standard deviation is large and the prior is therefore relatively weak, it does not bias results when the measurement signal -to-noise ratio is larger. Finally, the prior covariance matrix has zeros on the off-diagonals (i.e. no correlation between levels at the coarse resolution of the solution).*

The idea of calculating a prior from the analytic approach is something that some authors have used. It's important to distinguish between a prior and first guess; they may or may not be the same state vector, but they play different roles. The first guess is used to quickly start the iterations in a regime that is believed to be close to the correct answer; therefore, the analytic solution makes a very good first guess. However, the role of the prior is to encode any information that exists *outside* of the measurements (literally "prior" to the measurements), so using an alternate inversion of the same measurements doesn't satisfy that role. To explore the idea further, the prior uncertainty is what's critically important, even more than the prior state vector. If the prior covariance matrix correctly reflects the amount of knowledge outside the measurements (what you would know if you made no measurements), then even if the prior state vector is "borrowed" from the analytic retrieval, that state vector will only be perceptible in the OE results where the measurement signal is relatively weak, which are the same locations where the analytic retrieval is most unstable. So, I don't think there would be significant benefit to using the analytic retrieval for the prior state vector (but again, there is clear benefit to using it for the first guess). If you consider using the *uncertainty* for the analytic solution as a prior uncertainty, that would be a definite handicap, since the balance between the prior uncertainty and the measurement uncertainty is what drives OE to find the best smooth solution that fits the measurements. It would also disable the ability to calculate DOF of the measurements and the associated effective resolution.

4. Fig.5. What are the units? Are measurements normalized?
The figure labeling for Figs 2,5, 12, and 18 has been revised to say "arbitrary units". These values are the measurements or simulated measurements after relative calibration (i.e. multiplying by the gain ratios) and removing the range-squared dependence by multiplying by r^2. The figure annotations have also been changed to make this more explicit using the same symbols as in the equations, e.g. the quantity obtained for the molecular channel is $P_m r^2 g_p/g_m$. The question was specifically about Fig 5. Fig 5 has the same units as Fig 2, since it is just a difference (at each specific grid point), and likewise for Fig 18 and Fig 12. The overall scale is different between the simulation and the observed measurements, because the overall calibration constant is different. In the case of the simulation, since the overall scaling is 1, this quantity can be thought of as the attenuated backscatter.

5. Fig.11. I have difficulty to understand this figure. Why effective resolution in OE method increases in non-monotonic way? For grid of 50 m the uncertainty is the highest, but effective resolution is also high (700 m). Would be good to provide more explanations.
The effective resolution is not monotonic because there is a tradeoff or balance between the prior and the measurements in the cost function. At the finest grid scales, the retrieval pushes the limit of the information content of the measurements, and the prior begins to take effect. At 50 m, the prior plays a big role. The uncertainty is large because the prior uncertainty is large, and the effective resolution is also coarse because solutions at neighboring grid points are not independent (since they are all highly dependent on the prior information). This is the explanation given in the revision:

> *For OE, there is a minimum achievable effective resolution for extinction. This reflects the balance between the measurement uncertainty term and the prior uncertainty term in the cost function. Retrievals at finer grid scales don't achieve finer effective vertical resolution because the prior has a greater impact; the measurement uncertainty at these fine grid scales asymptotes to the prior uncertainty, and the resolution is coarse because solutions in neighboring grid points are not independent of the prior solution or of each other. At grid scales coarser than the one that leads to the minimum effective resolution, the usual trade-off between propagated uncertainty and vertical resolution prevails, with coarser resolutions associated with smaller propagated uncertainties.*

6. Fig.14f. At altitude of ~0.5 km uncertainty of OE is higher than for analytical solution. Can it be explained?

This single data point is within the clear air gap between two aerosol layers. In that region, the measurement information content is small and the uncertainty for both methods is quite large. I suspect the reason that the OE uncertainty is larger in Figure 14f at that point is that systematic uncertainties were very carefully included in the OE error analysis in this study, whereas the analytic uncertainties shown here are only the propagated measurement noise. This is mentioned elsewhere in the text, but in the revision, the figure captions for Figures 3 and 14 are also revised to reflect this. For example, the caption to Figure 14 (now Figure 11) now includes "(e)-(h) show the a posteriori uncertainty of the OE retrieval and the propagated random uncertainty for the analytic retrieval; these are also represented as error bars on the top panels."