Author (Daniel Letros) responses to the referee comments is shown in blue.

Our entire team would also like to thank you for your feedback. Your efforts have helped to improve our work and we are appreciative!

The authors have put together a very impressive paper to provide a brief description of their instrument (the Aerosol Limb Imager, aka ALI) and its aerosol retrieval algorithm. Herein, the authors provide a thorough presentation of their algorithm and discuss some of its limitations in a comprehensive manner. The ALI instrument takes advantage of limb scatter measurements at multiple wavelengths for 2 polarization states. This information content enables the authors to retrieve aerosol radius and number density estimates as well as a rather rudimentary estimate of the distribution width, which is assumed (forced?) to be constant throughout the profile. I must reiterate that overall I am pleased with this paper. It is well written (up until sec- tion 4.3 where typos became more plentiful as did missing articles), logically organized, easily read/followed, and generally well supported with ample use of figures. I believe this paper will make a important contribution to the scientific literature, that it fits well within the scope of AMT, and should be published pending some revisions.

1. General comment: The authors used SAGE III/ISS, OMPS, and OSIRIS data to evaluate their derived extinctions, but nowhere do they tell the reader what profiles were used (e.g., profile or event number), do not tell the reader where these profiles were collected (e.g., lat/lon of the satellite observations), and they fail to inform the reader of when these profiles were collected. All of this information is necessary for interpreting these intercomparisons (some of this may explain the differences they observed between the satellite instruments and ALI). This information should be added before publication.

We have made a supplemental PDF which contains information tables and a map showing the coincidence of the ALI measurements with the other three instruments. We have also made references to the supplemental document in both the caption of Table 1 and elsewhere in the paper.

2. General: The authors failed to inform the reader of where ALI was sampling. The authors stated in section 3 that "ALI was situated on the balloon gondola and orientated such that when the gondola is flat and level, the highest lines of sight (top pixels of the ALI detector) would be horizontal and with tangent locations on the instrument itself " (emphasis added). The question is: where is the bulk of the scattering coming from? It's certainly not the "tangent point" (i.e., at the instrument itself). Maybe the distance is negligible, but the reader probably does not know that. This information should be included before publication.

We consider the location of the ALI scans to be at the mean tangent position of each ALI scan. We indicate this in the supplemental document that was made to address your comment above. In addition we have swapped the lat/lon information of Table 1 from being the position of the gondola (which is relatively unimportant information within the paper) to be that of the mean tangent positions of the measurements. The caption of Table 1 has also been updated to reflect this.

3. General: The authors miss an opportunity to compare their radius and number density estimates with those released by the SAGE team. If there were coincident OPC data available I would recommend they use that, but since the SAGE data is all that's available I must strongly recommend the authors include that data in their analysis before publication.

We have added this comparison with discussion to the paper.

4. General/Figures: The figures could benefit from being larger, which would enlarge the fontsize of the axes and make them easier to read. Please consider making this change.

Further efforts have been made to make the figures more readable and properly formatted in the final submission.

5. Page 6, line 122: The authors claim that the error is "...on the order of the square root of the DN values..." Maybe I am being too precise, but the FWHM in Fig. looks to be ≈40, which corresponds to a DN of 1600 (i.e., 40^2) whereas panels (a) and (b) indicate DN on the order of 8000. Would the authors please clarify their meaning here?

Your numbers are approximately correct. This statement was just meant to be a digestible one indicating that this methodology is not introducing significant error in the analysis. We have reworded it for this explicit intent:

Figure 4 shows an example of synthetic image construction compared against an actual ALI image. As this example shows, the synthetic image provides a faithful recreation of the actual ALI measurement. The error of this reproduction, as demonstrated by the histogram FWHM of 33 DN, is approximately 0.4% of the mean of the image signal.

6. Fig. 4: What is the black dot in panel (a)? That corresponds to the large white dot in panel (b), so I assume it is a bad pixel, but all other bad pixels are white. Would the authors please clarify?

White dots are indeed bad (non-ideal) pixels. The lack of any non-ideal pixel in (a) with respect to (b) comes from the level of calibration applied to the image. (a) is an ALI image of the integrating sphere which has only dark correction applied (i.e. only pixels which are identified as non-ideal under non-illuminated conditions are white). (b) is a synthetic reproduction of (a) which includes non-ideal pixels under both illuminated and non-illuminated conditions. The pixels identified as non-ideal under illumination (such as the block dot you mention) are determined from images like that of the one shown in (a).

The caption of Figure 4. does allude to this, but it is not as clear as it should be. We have adjust the wording:

Figure 4. Example comparison of synthetic image construction for ALI correction. (a) A single ALI image of the calibrated integrating sphere. AOTF is tuned to diffract 1450 nm, the LCR is off, and an exposure time of 0.450 seconds is used in this example. Dark correction has been applied. White dots in this image indicate the bad pixels as determined only by the dark correction. (b) A synthetic reproduction of the real image shown in the left plot constructed from the database of calibration coefficients. White (bad) pixels seen in this image are determined as pixels with non-ideal responses in both dark and illuminated conditions. These pixels are discarded. (c) A histogram of all (non-bad) pixel values after the synthetic (middle) image is subtracted from the real (left) image.

7. Page 7, lin 143: The authors refer to ALI's viewing geometry "...when the gondola is flat and level..." I have two questions regarding this:

(a) How often is the gondola level and flat? I assume there is some stabilization utilized, but it is never explicitly mentioned.

(b) Is this orientation monitored to allow correction? Was a correction applied? How does this variation impact the view geometry and the results?

The initial statement of "...when the gondola is flat and level..." is meant to indicate the fixed orientation of ALI mounted on the gondola. Of course, the gondola itself is not perfectly flat and level during flight and has its own time-varying attitude. To answer both (a) and (b) together: the gondola is pointed and its momentum can be transferred to the helium balloon giving stability to the instruments. The orientation of the gondola is tracked by IMUs which allow us to reconstruct each ALI line of sight of each image. In selecting the three scans used in the paper, attention was paid to the variation of the gondola during image exposures to minimize what would manifest as spatial blurring.

As alluded to in the paper, Scan 1 was almost completely stable in attitude over the exposures but the other two scans also have minimal changes (i.e. changes in orientation exceeding the uncertainty of the

IMUs). To provide an idea of the gondola stability you will find a plot of the flight reconstruction below. Scan 2 (taken at about the 7 hour mark in the attached plot) is the worse of the three used scans - although again we will emphasize it is still relatively stable with $< 0.1^\circ$ change in pitch over each image acquisition. While we did not present other retrievals of less stable scans in the paper, when studying those we binned detector rows together by tangent altitude as a strategy to accommodate the lost of spatial resolution of the measurements within the retrieval.

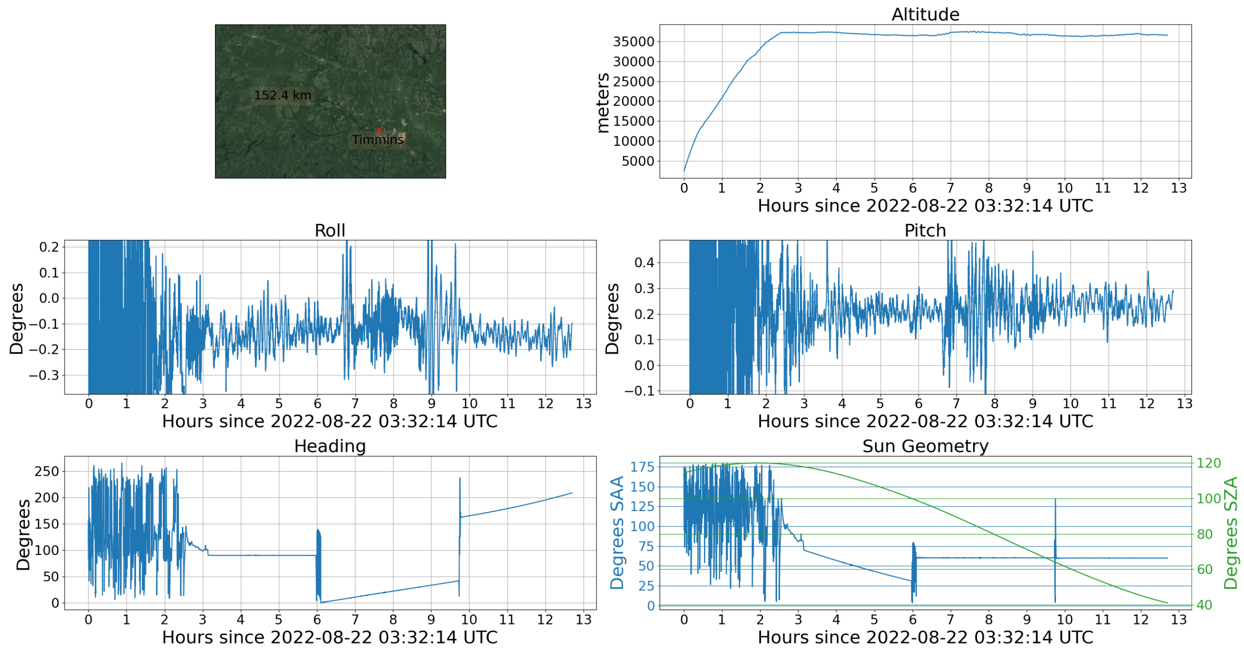


Figure 1: Reconstruction of ALI position and attitude on the flight gondola during the Timmins 2022 flight.

We refrain from discussion around the gondola attitude and flight reconstruction within the paper for brevity, as we hope to maintain scope around the retrieval algorithm and ALI science demonstration. However, for other readers we have now appended to the first paragraph of Section 3:

[...] At this point in the flight the gondola was steered to maintain a solar azimuth angle (SAA) of 60°. During the flight the gondola position and orientation is recorded which allows reconstruction of all the ALI lines of sight in each image.

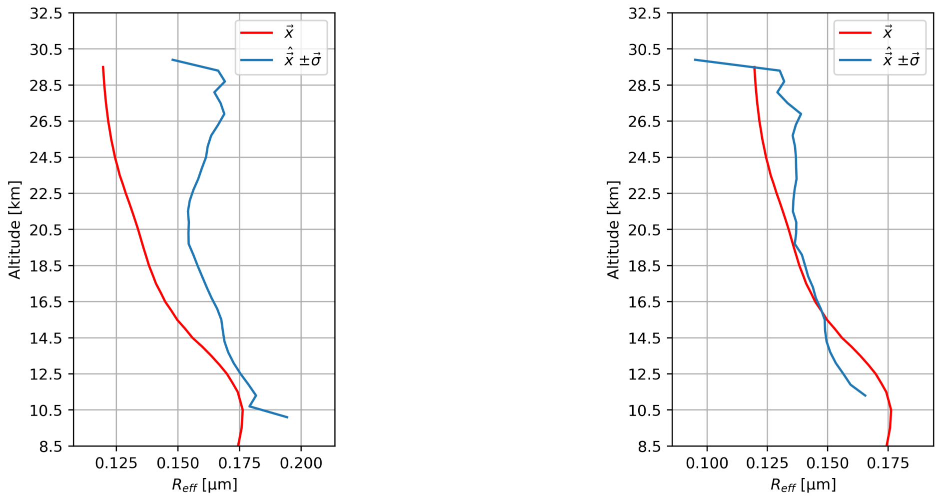
... and these in the third paragraph of Section 3:

[...] The balloon gondola was relatively stable for this scan compared to most others taken during the flight (gondola attitude within IMU error during exposures), [...]

[...] The other two scans we select present more difficult observation conditions both in terms of relative gondola stability (although there was still minimal attitude change during exposures, i.e $< 0.1^\circ$ change in pitch over each image acquisition) and observational conditions.

8. Fig. 12: I wonder if the authors sell themselves short on this figure. You cannot solve for bimodal distributions, but you may be able to reasonably infer the effective radius, which would resolve some of the "bimodal issues" the authors allude to throughout sections 5/6 as well as provide a more robust number for use in models. That said, the constant width value may skew re. I suggest the authors include another panel in this figure to evaluate re.

This is an excellent thought. We provide you the effective radius profiles of the retrievals in Fig 12 and Fig 13 below.



(a) Exercise of Figure 12 (geometry of Scan 3)

(b) Exercise of Figure 13 (geometry of Scan 1)

Figure 2: Aerosol effective radius of bimodal simulations.

In the specific case of simulation we do not think that they add any new insight or information beyond what is already shown. A core finding we are attempting to convey in the paper is that polarized limb measurements are more sensitive to aerosol particle size distributions. Unimodal treatment of a bimodal distribution can be insufficient. To this end, failing to retrieve the correct extinction in the exercise of Figure 12 (using the geometry of Scan 3 where the atmospheric degree of polarization is higher than Scan 1) shows this shortcoming. Likewise, simulating the same aerosol state but in a less polarized atmosphere (geometry of Scan 1) in Figure 13 yields noticeable improvement. The effective radius profiles ultimately reflect the same improvement the extinction profiles do. We also attempt to convey that a similar looking improvement between Scan 1 and Scan 3 in the real retrievals (Fig 16 and Fig 18) is observed, although with a much less realistic state in Scan 3 as you note below.

9. Page 23, line 460: The authors state "In this simpler retrieval only N is adjusted in x such that we arrive at an aerosol extinction directly retrieved at 750 nm." As written, it sounds like the authors are iterating to match an extinction, when I think they are iterating to match the radiance (like OMPS and OSIRIS). Would you please clarify?

We have reworded this sentence (along with other minor rewording later in the paragraph) in an attempt to be more clear :

In this simpler retrieval only N is retrieved in the state vector $\vec{x}$, with only the 750 nm radiance constructing $\vec{y}$.

10. Page 24, line 465: The authors state "Retrieving extinction at only 750 nm yields respectable..." (emphasis added). "Respectable" is ambiguous, please quantify and clarify. Addressed in 12.

11. Page 24, line 469: The authors state "...shows a fairly ideal retrieval." What is meant by "fairly ideal"? I have no idea what is meant. Can the authors be quantitative or provide a metric for the reader to gauge "idealness"? Addressed in 12.

12. Page 24: line 470: The authors state "...the retrieved extinction well represents the ex tinction profile of all three comparison instruments." Again, "well represents" is ambiguous. The figures in reference (15, 16, 17) are plotted on log scales, which makes quantitative evaluation challenging. Readers would greatly benefit from a percent difference plot (or a ratio plot where the 3 comparison instruments are divided by ALI's extinction). This would convey a wealth of information to the reader. Would the authors please include these plots?

Addressing points 10, 11, and 12 together: We have added percentage difference plots to the results and added wording to quantify such statements.

13. Fig. 18: I think the authors are overly optimistic in the interpretation of panels (a) and (b). There is no way the number density is so large (it's at least an order of magnitude too high, even after a major eruption) and the sensitivity of their instrument (at the designated wavelengths) to particles with radius of 0.04 µm is questionable. I would suggest that if the authors want readers to take panel (b) seriously then they should provide more support (maybe show scattering intensity at the current scattering angle for the wavelengths in question as a function of particle size?).

We do not expect the state of this retrieval to be taken seriously in an absolute sense - especially so within scope of comparison to the Scan 1 retrieval and the newly added SAGE particle size comparison. In line 492 I state "We speculate that this is another manifestation of the retrieval trying to optimize a unimodal distribution to match the polarized $\vec{y}$ produced by a more complicated aerosol population." indicating we don't take it seriously ourselves. The discussion surrounding this part of the paper is attempting to highlight the positive aspect that the retrieved profiles are similar in shape to Scan 1, and further demonstrating this polarization point.

14. Page 29, line 522: Regarding the sentence starting with "However, the disagreement of these two...": The authors put all the blame on a bimodal distribution, which may or may not be the case. However, I don't think this is supported and I am unsure of what the authors are trying to communicate within the last 2 sentences of this paragraph. Are they suggesting that the instrument is seeing different atmospheres in the various scans and scan 2 observed aerosol with a bimodal distribution? Are they suggesting that the instrument is seeing the same atmosphere, but the profiles are different because a bimodal distribution has more impact at some scattering angles than others? I don't know. Would the authors please clarify?

We have clarified the wording:

[...]Scan 2 and Scan 3 showed an overestimation of aerosol extinction with respect to SAGE III, OMPS, and OSIRIS. However, the disagreement of these two scans can potentially be explained by the affect of using a unimodal distribution to represent more complex aerosol under polarized limb measurement conditions. We speculate that the same underlying retrieval behaviour seen between the simulated exercises of Fig. 12 and Fig. 13 may be occurring in the real retrievals as well. Supporting this statement is the relative improvement in the quality Scan 1 exhibits which is also replicated in simulation under this case.