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

We would like to offer our appreciation for your efforts and feedback. While some of your notes may be outside of what we consider the scope of the paper to be, we would like to say again that they are not bad suggestions. Much of what you suggest is of significant interest to us as well and it is simply a matter of time and resources to investigate them. Regardless, we feel your feedback has helped greatly to make this paper more presentable. Thank you!

Referee report to the "The Aerosol Limb Imager: Multi-spectral Polarimetric Observations of Stratospheric Aerosol" manuscript by Daniel Letros et al.

The manuscript describes a retrieval algorithm for the new Aerosol Limb Imager (ALI) instrument and presents some results from the synthetic retrievals and from three example measurements made by the instrument. Although the measurement concept of ALI is similar to that used by the upcoming ALTIUS mission of ESA, it offers an unique feature of measuring the polarization state of the limb-scatter radiance. Authors did a great job showing how this feature of ALI can be used to detect contamination by clouds, which has always been an issue for limb-scatter aerosol retrievals. Less impressive are, however, the results from the aerosol retrieval itself. Here, I got an impression that the polarization-sensitive measurements make the instrument useless for the aerosol retrieval. This is not an issue for the scientific significance of the paper but the authors, especially PIs of the project, should think about if they really want to provide this impression to the scientific community. Unfortunately, the presentation of the results is elaborated quite poor and needs to be improved. To improve the message of the paper authors need to quantify the required conditions for the retrieval of reasonable aerosol extinction coefficient profiles from real measurements. A careful proof-read of the paper is needed to correct typos, extra or missing words etc.

There are two points we would like to generally highlight:

First, we understand there are some issues with the direct comparison of ALI retrievals against the results of SAGE III/ISS, OMPS, and OSIRIS. Principally, ALI retrieved extinction is biased high with respect to the other three instruments. We do not consider that such bias is unreasonable given our measurement platform of a high-altitude balloon (altitude of 36-37 km), which imposes analysis difficulties that are significantly less relevant to a satellite platform in orbit. We also do not think it invalidates the core of our demonstration.

However, we did fail in properly contextualizing and acknowledging this as the paper was. As a broad response to your feedback we have added more discussion (primarily around Fig 15 of the paper) which attempts to provide this context and acknowledgement. This also relates to some of your other feedback given below that we address further in those responses.

Secondly, there are a number of comments below that we would like to give a broad address to. The intended scope of this paper is to provide demonstration of the ALI instrument concept from a polarized limb measurement level (i.e. the Timmins 2022 campaign) through the current analysis algorithms we have developed arriving at a level 2 product of aerosol and cloud height. This demonstration culminates in the comparison of ALI results with SAGE III/ISS, OMPS, and OSIRIS for three example sets of ALI data. We provide simulated exercises to contextualize and support this comparison, and consider that these three scans well represent both the efficacy and limitations of our approach at the current state.

We feel that some of the comments below - while not bad suggests to give - tend to expand the already lengthy paper outside of what is required for this intended scope. We still respond to all comments below, but if we include an "out of scope" comment in our response we simply mean to say that we consider what is requested to be beyond the scope we have defined here.

General comments

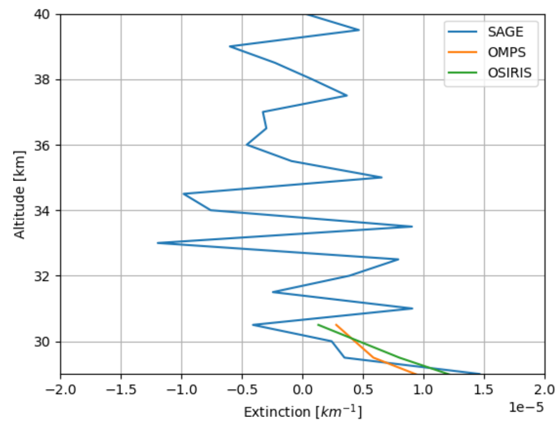
- Date and time of the ALI measurements and those of the collocated reference measurements are not provided. This make impossible to understand which aerosol conditions were investigated. Furthermore, no information about collocation criteria is given.

  - We have adjusted Table 1 to include more relevant information to the ALI retrievals rather than the flight gondola. Additionally, we have added a supplemental document which contains further information about the collection and coincidence of the ALI measurements to the other three instruments.

- Simulation results do not look representative enough. The aerosol extinction above 30 km was set to zero and no attempts were made to check what happens if a realistic aerosol distribution at high altitudes is included. It is not clear for which surface albedo the simulations are done and what happens if albedo changes. How the estimations of the surface albedo are affected by the presence of the aerosol above 30 km? Are the simulation results remain the same if an aerosol profile for different aerosol loading conditions is used? Logarithmic plots and absence of relative difference plots make evaluation of the retrieval quality nearly impossible. If I understand it correctly, the synthetic retrievals were done without adding measurement noise to the simulated spectra. How well the retrieval chain works if measurement noise is added (I mean here using noisy simulated spectra, not only adding the noise covariance matrix into the retrieval)? How the aerosol parameters for the exercise with bimodal aerosol PSD were set, where comes the information about the used parameters from? Are they realistic? Some quantification of the results in the case of the bimodal PSD is needed, i.e. some realistic bimodal distributions corresponding to different aerosol loading conditions have to be found and used for simulations. Comparison of the mode radius and width from the unimodal and bimodal distributions does not make much sense. A comparison of the effective radii might be more useful.

  We will address different points here individually:

  - Simulation results do not look representative enough. The aerosol extinction above 30 km was set to zero and no attempts were made to check what happens if a realistic aerosol distribution at high altitudes is included.
  - How the estimations of the surface albedo are affected by the presence of the aerosol above 30 km?
    * We are unsure what you mean by a realistic aerosol distribution in this context. It is our understanding that it is realistic to consider sulphate aerosol zero at (approximately) 30 km and above. Of course there is some latitude dependence, but it is at about this altitude that the sulphate aerosol will evaporate into gas. As we also understand it, it is typical for Lidar aerosol measurements to use altitudes of 30 km and above as a Rayleigh signal for calibration. For the purposes of the paper, this assumption is supported by the observations of SAGE, OMPS, and OSIRIS which we include below. We do understand that it is not strictly correct assumption to make **- and this is an aspect of our address around Fig 15. in the paper which pertains to this and other limitations of our approach -** but this is an assumption we must make from a balloon platform which does not have access to even higher altitudes to normalize from.



High altitude aerosol extinction. Coincident profiles of Scan 1 used.

- * To elaborate further on the normalization: when normalizing the measurement vectors, the aerosol one retrieves will be done with respect to the aerosol signal within the normalization altitudes. Again, ideally this aerosol signal would be zero and we make this assumption implicit in the paper due to limitations of the balloon platform. If this assumption is incorrect then the retrieved aerosol profile will be biased low by some proportion to the signal in the normalization.
  * Robustly exploring this in simulation for different aerosol loading and/or albedo is computationally non-trivial, and we consider it out of scope of this paper given what we already present.
- It is not clear for which surface albedo the simulations are done and what happens if albedo changes.
  * For the simulations, unless explicitly stated (i.e. the albedo estimation discussion), the albedo is held the same in both the simulated atmosphere of the true state and the forward model of the retrieval. This removes it as a factor in these simulations. In general, if the albedo between the true and retrieval atmospheres is incorrect then the aerosol will be biased. If the albedo is too low in the retrieval versus the true state atmosphere, then the missing scattered signal will be attributed to aerosol by the retrieval and the retrieved aerosol profile will be biased high. If the albedo is too high the opposite happens. Again, we consider robustly exploring this in simulation out of the paper's scope - **but it is a part of our new discussion around Fig 15. of the paper**.
  * For the flight retrievals the albedo estimation is an effort meant to align the forward model with the up-welling signal seen by ALI. However, the high-altitude normalization is the main strategy to marginalize the effect of albedo.
- Are the simulation results remain the same if an aerosol profile for different aerosol loading conditions is used?
  * Yes, the results of the simulation reflect well the performance of the retrieval even with different aerosol loading. The caveat to this statement is if true state conditions and a-priori conditions differ significantly - i.e. the underlying a-priori information given to the Kalman filter is no longer appropriate to reach the true state solution. This is a broad topic and a limitation of any Kalman filter applied to any problem (not just atmospheric inversions). As such the exercise of prototyping a Kalman filter and exploring it in simulation to yield appropriate performance is always necessary.
  * We have added particle size comparison to SAGE III/ISS, and with that we also now discuss the influence of a-priori selection to a degree. However, presenting different aerosol loading in simulation (true state or the a-priori) is beyond the scope of this paper.

– Logarithmic plots and absence of relative difference plots make evaluation of the retrieval quality nearly impossible.

  ∗ We have added relative difference plots to the results now.

– If I understand it correctly, the synthetic retrievals were done without adding measurement noise to the simulated spectra. How well the retrieval chain works if measurement noise is added (I mean here using noisy simulated spectra, not only adding the noise covariance matrix into the retrieval)?

  ∗ We believe there is confusion here. The synthetic retrievals were done with measurement noise applied to the simulated observations constructing the measurement vectors, and not just in the noise covariance matrix. To avoid any further confusion we have now noted this explicitly at the beginning of Section 4.3.1.

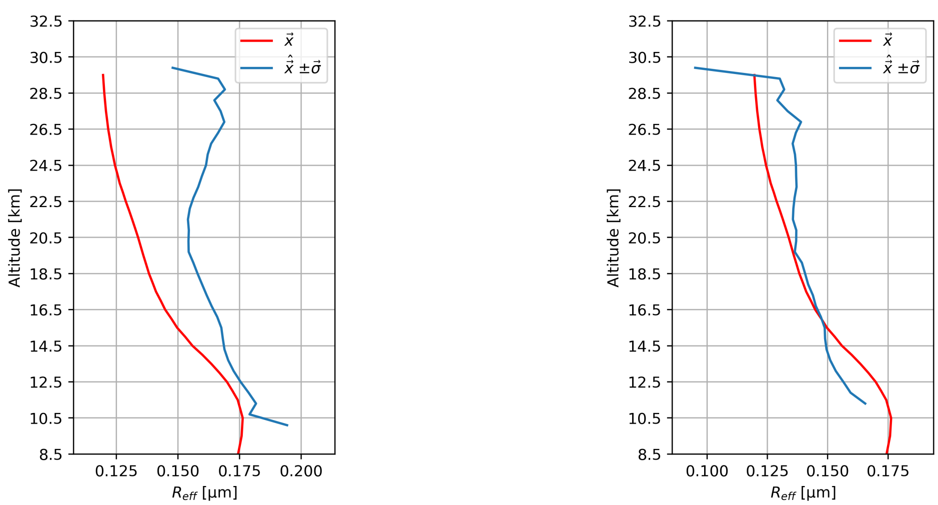
– How the aerosol parameters for the exercise with bimodal aerosol PSD were set, where comes the information about the used parameters from? Are they realistic? Some quantification of the results in the case of the bimodal PSD is needed, i.e. some realistic bimodal distributions corresponding to different aerosol loading conditions have to be found and used for simulations.

  ∗ We do not have any particular reference for the bimodal aerosol distributions. We simply altered the sizes ourselves by increasing the radius with a correspond decrease in width. This is to approximate larger aerosols coagulating with smaller ones. Beyond that we simply made the two profiles with distinct particle sizes in altitude with respect to each other. Since these parameters do not vary significantly from the a-priori sizes we use (0.08 microns and 1.6 width which is also used by OMPS-LP) we consider them realistic.

  ∗ We agree that a robust study of the impact bimodal aerosol loading has on polarized unimodal aerosol retrievals (beyond what we show) would be very interesting. However this study will be *very* time intensive and computationally expensive. For now we consider it to be outside the scope of the paper.

– Comparison of the mode radius and width from the unimodal and bimodal distributions does not make much sense. A comparison of the effective radii might be more useful.

  ∗ This is a very good thought, and we show the effective radius of the bimodal simulations below. However, in the specific case of simulation (where the true state is known) we do not think that they add any new insight or information beyond what we already show.



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

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

Figure 1: Aerosol effective radius of bimodal simulations.

        * To elaborate more: 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.

- Comparison with measurement data is difficult to evaluate without having the relative difference plots. It is not clear for which aerosol conditions the comparison with reference data was made, this information is extremely difficult to derive from the logarithmic plots. The overall conclusion from the comparison seems to be that only one of the three measurements produces reasonable results although data from both OSIRIS and OMPS-LP work well in all three cases. Is it a general issue of the applied technique? Is the technique then useful at all? A more detailed discussion of the usability of ALI results needs to be done at this point.
    - As we noted above, we have now added some relative difference plots for the extinction comparison. To better discusses the usability of ALI we have augmented the conclusion with the hope of addressing this concern.

Detailed comments

- Line 6 and 34: At this point it is unclear what the term "scalar width" means. Clarified.
- Introduction lacks some information about how the ALI instrument is ranged with respect to the previous, present and planned space-borne aerosol instruments.
    - The version of ALI in the paper is of course a balloon instrument, and while the ALI concept is planned to become a space-borne instrument, that version is still in development. Perhaps we misunderstand the intent of this comment, but we are unsure how to appropriately range ALI with respect to space-borne aerosol instruments beyond what we do for these reasons.
- Section 2: Some technical data of the instrument need to be provided, e.g. vertical/horizontal resolution and sampling, FOV range, typical exposure time etc.
    - We have added some more detail to the beginning of Section 2:
        * ALI is designed to image the atmospheric limb from a float altitude above 35 km. It has a 6° full field of view in the vertical dimension and a 4.8° in the horizontal. To image the limb with this field of view ALI is tilted down about 3° so the top lines of sight are flat and level with the (idealized) surface of the Earth. This external field of view is mapped to a 640×512 pixel detector which, including the instrument point spread function, yields a 0.06° angular resolution for each pixel in both dimensions of the external field of view. This translates to a tangent altitude resolution of < 100 m. Exposure times vary on ALI configuration (i.e. imaged wavelength) and the solar scattering geometry, but it is typically a few hundred milliseconds.
- Lines 51 - 52: "The SPS also contains one linear polarizer after the LCR and another after the AOTF to further refine the polarized image" – There is a polarizer before the AOTF and AOTF itself passes one polarization direction through. Please explain shortly why an additional refinement is needed.
    - We have added a short explanation: [...] the AOTF to further refine the polarized image (Letros et al., 2024; Kozun et al., 2021). The refinement provided by the polarizers heavily attenuates any unwanted light which may otherwise be passed through the AOTF.

- Figure 1 is quite dark and depending on the monitor and illumination difficult to see. Please light up the dark parts of the figure. We have brighten up the figure by making the black objects a lighter gray.
- Figure 2, panel (b): the abbreviation "DE" is not defined in the caption. Corrected.
- Line 69: "The spectral bandpass of ALI for a tuned frequency is taken as the full-width half-max (FWHM) of the diffraction response." – it looks like, here, you use the "diffraction response" term to describe the same function as shown in the panel (b) of Fig. 2 an referred to as the "spectral response". In line 75 you write, however "spectral response is calculated by integrating the measured diffraction responses" which suggest different meaning of these two terms. Please clarify.
  - Yes we consider the diffraction response of the AOTF and the spectral response of ALI to be the same thing. In line 75 we write "the area of each spectral response is calculated by integrating the measured diffraction responses". We use diffraction response here just to emphasize it is an AOTF measurement, but we understand it may be confusing.
  - We have changed line 75 to read: [...] the area of each spectral response is calculated by integrating the measured responses in Fig. 2. (b) [...]
- Figure 3 caption: "LCR on state" – here and throughout the text, it would be less confusing is you wrote "on" in the quotation marks. The same is for "off".
  - We have actually received the exact opposite feedback in the past from others. Our initial instinct was to write it as you suggest, but we think without quotations is also just fine. Suspecting we can not make everyone happy on this one, we will elect to keep it as currently written.
- Line 99: please give some details to explain what the "calibrated broadband integrating sphere of known (randomly polarized) spectrum" is.
  - We are not sure how to provide more detail which is both relevant and appropriate here to satisfy your comment. An integrating sphere is a common piece of optical equipment used in calibrations. We are simply saying we are using one which has been calibrated, and that we know well the spectral output of the source.
- Line 100: "Furthermore, since they reproduce the spatially flat and full-field conditions of the integrating sphere, they can be used to relate the ALI measurements to the external source each pixel measured."– Does "spatially flat" means homogeneous of a flat shape of the source? It is not quite clear how to get a response to an inhomogeneous source signal if you have a response to a homogeneous full-field signal. Please explain. Maybe I misinterpret what you want to say with this sentence, in this case please rephrase for more clarity.
  - Yes the integrating sphere is providing a flat/uniform source for ALI to image during calibrations. This sphere also illuminates all field angles of ALI just as the atmosphere does. Each pixel of the detector effectively just counts how many photons land on it, and since we know the intensity of the integrating sphere we can relate the raw detector images to units of radiance on a pixel by pixel basis. As for the response of an inhomogeneous source (like the atmosphere) that comes down to a matter of the instrument point spread function which defines our spatial resolution. We cannot spatially sample the atmosphere better than our point spread function allows, and within this resolution the atmosphere is effectively a homogeneous scene. More details about the relation of ALI image intensity to the atmospheric radiance can be found in (Letros et al., 2024) which includes polarized aspects of this.
- Line 128: Are there any investigations of the temporal stability of the obtained calibration parameters and possible dependence on the ambient temperature and pressure, strength of the illumination etc. ?
  - Temporally yes. Calibrations are performed at different instances over time. For example, we do calibration procedures within the lab both before and after events like a flight campaign. Not only does this tell us temporal stability on long time scales, but also if something happened to the instrument during flight. We find no concern in temporally stability.

- In regards to temperature, much design effort was given to monitoring and controlling the temperature of the instrument during operations. Particularly to the SPS. Calibrations are performed around nominal conditions which are well maintained during flight except for ascent conditions through the tropopause - and we do not attempt measurements in conditions like this. However given this approach, with the exception of the detector, temperature is not considered a variable in image processing
  - The gain of our detector was explored as part of the conversion image calibration. Including adjusting the source brightness of the integrating sphere. From lab measurements the gain is very linear within our operating parameters (exposure times, AOTF configuration, etc). So we do not take any variation of detector response with respect to source strength into account.
  - Pressure and other aspects of similar granularity being to go beyond the scope of the calibrations we did.

- Figure 3 caption: "White dots in this image indicate the bad pixels as determined by the dark correction." – I do not see any white dots in panel (a), what do the black dots mean?
  - We think you mean Figure 4 here (the figure showing the synthetic reproduction of an ALI image). There are white dots in Figure (a), but they are subtle. They are bad pixels as determined by dark correction only. The black dots in (a) which are white in (b) are pixels with poor response to illumination and are not yet flagged bad at the stage of processing seen in (a). We have adjusted the caption of this figure to be more clear.

- Figure 5 caption: The sentence "The radiance profiles of (b, c, e, f, h, i) are constructed by following Section 2.3 to convert images into photons/s/cm2 /sr from DN. Then following Section 2.1 to obtain units of photons/s/cm2 /sr/nm." should be moved to the main text. Agreed and done.

- Figure 5 caption: "the solar irradiance produced by a radiative transfer forward model" – solar irradiance is not produced by radiative transfer models.
  - Yes we can be more clear on the wording. We have changed "produced by a" to "used by the".

- Line 206: "a regularization matrix in place of S_a^{-1} " – S_a^{-1} is also a regularization matrix, do you mean Twomey-Tikhonov regularization matrix, i.e. smoothing constraint here?
  - We have clarified the wording to say we are not using a Twomey-Tikhonov regularization matrix.

- Line 208: "measurement and state vectors of large dynamic range which tends to produce ill-conditioned inversions." – the ill conditioning of the inverse problem is not caused by the large dynamic range of the measurement and state vectors. Its reasons are rather a dense vertical sampling, insensitivity of the retrieval to certain vertical ranges and correlation between different parameters.
  - We completely agree that the reasons you list can cause or contribute to ill-conditioning, but we think dynamic range can lead to ill-condition the retrieval too. For instance, constructing $\mathbf{S}_\epsilon$ as we describe in the paper (diagonal matrix of the measurement noise) with notable difference in magnitudes between wavelengths can ill-condition the $\mathbf{S}_\epsilon$ matrix. This leads to ill-condition of the gain matrix and averaging kernel.

- Line 214 - 217: In most previous aerosol retrievals the albedo and aerosol retrieval were run alternatively within an iterative process allowing the retrieved albedo to adjust to the retrieved aerosol profile and vise versa. In this retrieval, the albedo is retrieved only once before the aerosol retrieval. Additional investigations need to be done to show that the retrieved albedo does not depend significantly on the a priori aerosol profile used for the albedo retrieval.
  - We addressed the choice of this algorithm in the paper (first paragraph of Section 4.1). It is to remove albedo from the retrieval optimization and reduce the solution space the retrieval needed to consider. As for your call to additional investigation, we refer to our address in the general comments above. However, we will note again that our new discussion of our limitations (included around Fig. 15) touches upon the albedo.

- Line 221 - 222: I do not expect that the influence of aerosol at 33 – 34 km is minimal. This statement has to be proven using measurement data, e.g. GloSSAC.
  - We refer to our address in the general comments above.
- Sect. 4.1 please provide the aerosol profile used for albedo retrievals. Dependence of the albedo retrieval on the assumed aerosol profile needs to be investigated.
  - We refer to our address in the general comments above.
- Lines 245 - 246: "Therefore, we also present an approach to retrieve the Stokes parameters of the atmosphere in the Stokes basis of ALI ..." – Please explain what "the Stokes basis of ALI" means.
  - The Stokes basis is of course the coordinate frame in which the Stokes vector is being described in. By Stokes basis of ALI, we simply mean we retrieve the Stokes parameters as they look to the ALI instrument. If for instance the flight gondola is rolled, the Stokes basis of ALI will be different than one aligned to the Earth's horizon (or some other frame of reference).
- Eq. (7): What happened to U and V components. Were they just neglected? Please clarify.
  - Equation 5-7 note proper definitions, as well as approximations of I, Q, and the DoP P. The approximations can be built directly from ALI measurements (without the more sophisticated retrieval approach). So yes, $\tilde{P}(\lambda)$ in equation 7 ignores U and V in the approximation.
- Lines 267: "...which directly correspond to tangent altitudes at the time the observation was taken" – What do you want to highlight with "at the time the observation was taken", are the tangent heights measured not at the same time?
  - This sentence is related to equations 8 and 11 which indicate elements of the vectors by detector pixel. With changing orientation of the gondola, any particular pixel may measure a different tangent point and altitude over time. This sentence is indicating that this relation is known, and the retrieved value of the n^{th} pixel can be related to altitude.
- Lines 300 - 302: "In this example Θ received little to no action by the inversion to adjust it form the a-priori state. It is sensible to simply not include Θ as a property in x and just rely on the a-priori values in the forward modeling." – Is it a common behavior or does it just happen occasionally?
  - Yes, in simulation we found this to be a very common behaviour.
- Line 310: "... we convolve the smoothed DoP with a central difference impulse response.." – please explain what is the "a central difference impulse response" or give a reference.
  - Here we are just describing taking a numerical derivative with the central difference method. We do this through a convolution, and in this vernacular it is called the impulse response.
- Fig. 7: Please provide an illustration what happens if there is a strong aerosol level with larger particles in place of the cloud layer.
  - Ultimately this is a point of future work for us, and we feel including this in the paper at this time will quickly lead out of the intended scope. Fig 7. depicts the cloud detection method of ALI in simulation. This, as presented in scope, encompass the retrieval of a quantifiable DoP used to identify altitudes of depolarization. The second paragraph of Section 4.2 states our intended scope and alludes to future interest.
  - To elaborate more: depolarization is associated with clouds because of enhanced scattering and anisotropy of the scatters (i.e. ice particles). In this work we model aerosols as isotropic sulphate droplets (which is a standard way to model stratospheric aerosol). We consider that the natural polarized distinction between the isotropic and anisotropic scattering gives a significant advantage to cloud detection. However, meaningful and robust investigation into the efficacy of cloud detection in the presence of larger aerosols will require modelling a slew of relative ice and aerosol

sizes/concentrations at different solar geometries/wavelengths (i.e. discussion of Fig 9) to study. We simply haven't devoted the resources to do this yet. That said, within the particle sizes we explored we haven't found ambiguity with the cloud properties we have explored.

- Line 333: "At each scan, the true state of the atmosphere includes GloSSAC aerosol but unlike the exercise of Fig. 7 and Fig. 8 no ice layer is included." – please show the aerosol profile used in the retrieval (if you like in the Supplement) and prove the results are the same for a different aerosol loading scenario.
  - The aerosol extinction we use is the same profile we show in Figure 10, but we don't consider the aerosol a primary driver of this effect. This manifests from the relative balance of horizontally and vertically polarized light within the atmosphere, which is largely driven by the scattering angle. The purpose behind the exercise is to demonstrate the failure of the DoP retrieval as a function of the ALI Mueller matrix. We believe the contrast between the top and bottom rows of this figure accomplish that as is.
- Sect. 4.2.4: It would be interesting to have a larger statistics about the working and non-working DoP retrievals.
  - We assume by larger statistics you mean more case studies (i.e. different solar geometries and wavelengths). We do agree and this is on-going work. For the paper though we consider this out of scope.
- Line 349: "retrieved along side a scalar width" – you still haven't properly defined what scalar width means.
  - Addressed in your first detailed comment.
- Line 369: "as the finest resolution A produced" – What does "A" mean here?
  - The averaging kernel **A**. It is bold in line 369 and defined in line 187. Although we see now the confusion it may cause. We have changed it to simply call it by name here.
- Lines 368 - 370: Resolution might be limited by the forward model grid but not determined by it. It is mainly determined by the instrument FOV and sampling and also applied regularization. What you mean here is the sampling.
  - We think we agree fundamentally, but perhaps there is confusion over wording. We are defining the altitude resolution in the forward model by determining the finest resolution which no longer increases the quality of the averaging kernel - i.e. if we increase the model grid resolution from 600 m to say 400 m, the resolution indicated by the averaging kernel will still be 600 m. In this case the measurement vectors we are using (for reasons like you note) lack the information to solve the state at a higher resolution. We have changed this sentence in the paper to try and avoid the confusion.
- Line 371: "we do not employ regularization" – Using Sa is also a regularization.
  - Addressed in comment of line 206.
- Lines 391 - 394: The description looks like the Levenberg-Marquard method, why do not you use the name, is the any significant difference?
  - Yes the dampening matrix **D** follows from the Levenberg-Marquard method. We have now included the explicit name when **D** is introduced.
- Sect. 4.3.1: Please show the GloSSAC aerosol profile used for the study. Why was only one profile used? Dependence on the true aerosol profile might be significant. Different aerosol loading conditions need to be investigated.
  - We refer to our address in the general comments above.

- Line 400: Please provide the illustration what happens if you do not set the aerosol profile above 30 km to zero.
  - We refer to our address in the general comments above.
- Figure 10: Is it correct that the true surface albedo is not changed here?
  - Yes, the albedo matches between the forward model and the true state.
- Line 416: What happens if a priori and true widths are the same but the width is included into the retrieval?
  - Good question. Using the a-priori covariance of 0.0001 for the scalar width tells the retrieval there is a 68.2% confidence (one sigma) that the true state width matches the the a-priori value width within 0.01. Because of this statistical freedom, the retrieval will end up moving the width off of the (correct) a-priori value. However, this movement is minimal and less than what we show in Figure 11. Of course one can tighten the covariance to reduce this effect, but these details are aspects of prototyping and tuning the retrieval algorithm.
- Figs. 10 - 13, 15, and 16 must include relative differences. Added.
- Line 433: "We wish to emphasize that the limb measurements of ALI are polarized, and we speculate that this polarized content contains useful information about the aerosol phase scattering matrices - which is of course influenced by the particle sizes" - also non-polarized measurements contain information on the particle sizes.
  - Yes, size information in general is available (although still challenging to retrieve) with measurements over wavelengths and/or scattering angles. Here we wish to emphasizes that the polarization information itself contains aerosol size information in addition to that obtained over wavelength and scatting angle.
- Sect. 4.3.1: Authors found a strong dependence of the results on the presence of the second mode for highly polarized measurements, the question is if polarized measurements are really an advantage or a drawback. This should be discussed in more details in the text.
  - We have attempted to make more clear the message we wish to deliver in the text. However to briefly clarify here as well: a purpose of the bimodal simulations was to showcase the failure of the retrieved measurement vectors reproducing the measurements under a higher degree of polarization. Similar behaviour is seen in the real retrievals as well. We would argue that this failure is a (at least potential) indication of the presence of more complex sizes - which itself is size information.
- Lines 448 - 451: "With respect to the simulations done in the exercise surrounding Fig. 9, we find almost all behavior regarding Fig. 14 to be expected including: the relative balance between horizontal and vertical polarizations for all three scans, the spectral regions expected to fail given the polarimetric response of ALI in Scan 1 and Scan 3, and the relative magnitude of the DoP for Scan 1 and Scan 2." – this discussion is difficult to follow, please write in more details what exactly you expect from the discussion around Fig. 9 and what we see in this respect in Fig. 14.
  - We have removed the reference to Fig 9. and it now reads "With respect to the simulations". We think this reference is where the confusion may arise. We are simply trying to say that polarized aspects of the atmosphere in all three scans are expected from forward modelled simulations - except the magnitude of the DoP seen in scan 3.
- Lines 448 - 451: "The simulation of this geometry done for Fig. 9 produced a true state DoP approximately ranging between 0.3 - 0.4" – please provide corresponding plots, if you prefer they can be in the Supplement.

– We have again removed the reference to Fig 9 to perhaps relieve confusion. We do not intend to include the plot of the DoP you request in the paper or supplement as we believe it adds nothing of value above what we state in text - but we do show it here immediately below. The significant difference we find between Fig 14 (c) and this plot is the magnitude of the DoP. We expect it to be in the range of 0.3-0.4 as the plot here shows.

– As Fig 7 and the Fig 14 (a) and (b) show, cloud (which is not in the forward modelled atmosphere for the DoP below) should yield DoP much lower then 0.3, and aerosol loading (which is included in the forward modelled atmosphere for the DoP below) should range roughly in the 0.3-0.4 range we see. The lump at about 18 km you see below comes from the aerosol loading (see the true state aerosol of Figure 10).

– As we say in the paper, we have no reason to consider the result of Fig 14 (c) to be wrong. We do not know what is yielding it yet. We suspect as the future work of exploring the DoP under different conditions (my response to your comment of Fig. 7. above) may yield more insight.

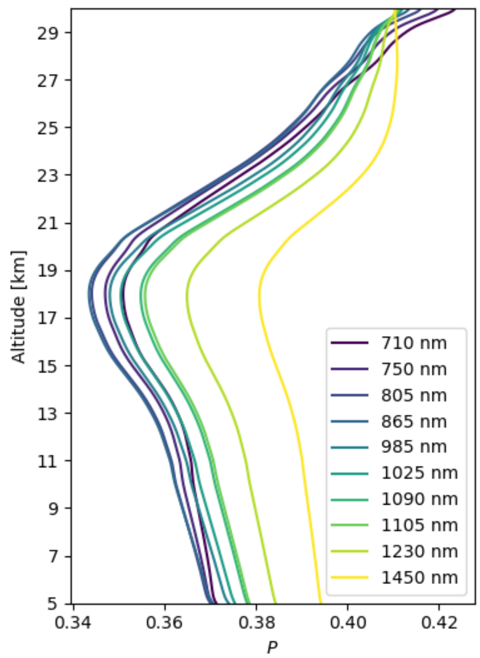


Figure 2: True forward modelled DoP of Scan 3 (exercise related to Fig 9)

- Line 473: "... of interest is the profile of r which indicates a layer of larger particles at approximately 22.5 km..." – please compare this results with SAGE III data and make a statement about the agreement. We have added a comparison to SAGE III particle size.

- Lines 485 - 488: A large disagreement for ALI and a good agreement for other instruments means, in my opinion, that the instrument concept where the polarized measurements are used is not optimal for the retrieval of the aerosol extinction coefficient. Please discuss this issue.

  – We acknowledge that ALI is an outlier to the other instruments. Our new discussion of limitations around Fig. 15 of the paper now addresses this more directly. However, the best extinction coefficient is built with accurate particle size. With what we have shown in the paper (and as you noted in the comment of Sect. 4.3.1 above), we have reason so think that the polarization includes useful size information which ALI can probe. Under our current unimodal approach, this has limiting factors which may (partially) explain why we are the outlier, but this is a point of ongoing work for improvement. We have attempted to convey this message better in the overall text as well.

Technical corrections

All technical corrections below are accepted and changed in the paper. Thank you for your diligence in spotting them!

- Line 33 and throughout the text: “dependant” → “dependent”
- Figure 1 caption: “comprises of one off-axis” → “comprises one off-axis”
- Figure 2 caption: “shown in black” → “is shown in black”
- Figure 3 caption: “response of ALI (Letros et al., 2024) shown as” → “response of ALI (Letros et al., 2024) is shown as”
- Line 117: “be used to constructed” →“ be used to construct”
- Line 128: “conversion to from DN into” → “conversion from DN into”
- Line 170: “Here we discuses the concepts” → “Here we discus the concepts”
- Line 175: “and marks the lower altitude limit to retrieve” → “and mark the lower altitude limit to retrieve”
- Line 347: “to emphasise that the in the context” → “to emphasise that in the context”
- Line 347: “criteria we set is to retrieve” → “criterion we set is to retrieve”
- Line 349: “their is” → “There is”
- Line 364: “ceiling on of the radiance” → “ceiling on the radiance”
- Line 396: “summary of retrieval result” → “summary of retrieval results”
- Line 407: “by retrieving a N and r profile” → “by retrieving N and r profiles”
- Line 413: “by the biased caused” - do you mean “by the biases caused”?
- Line 417: “shows the behaviour” → “show the behaviour”
- Line 515: “but we this is a point” → “but this is a point”