General comments on Letros et al. [2025]:

This is a well-executed paper that presents analysis of data taken by the ALI instrument, which records polarized radiance spectra. The value of these measurements is demonstrated through both sensitivity studies and analysis of selected ALI data. A particularly useful capability provided by radiance information is the ability to distinguish between clouds and aerosols along the line of sight.

I have few criticisms of the work presented, and particularly appreciate the degree of testing presented for the algorithm approach taken (which is clearly designed to test its limits, rather than primarily to show its performance in the best possible light). Error analysis is thoughtful and thorough. Some structural suggestions and clarification questions follow, but I have a positive opinion of the paper, and would be glad to see it published.

Detailed comments:

Abstract, line 4: "Tangent altitudes which have signal contaminated by clouds" should be defined more clearly. Throughout the paper, similar wording is used, and the authors' interpretation of this language appears to be consistent: This language is meant to describe tangent altitudes for which a cloud appears *along the line of sight*. But that's not the only possible interpretation: For example, limb scattering observations can (and frequently are) "contaminated" by upwelling radiation from a broad area below the line of sight, for which the locations and properties of the clouds are not well known. That is a distinct problem from interpreting measurements with clouds along the line of sight, and it isn't the focus of this paper, which the abstract should make clearer.

Abstract, lines 14-15: The abstract ends by declaring "good agreement" between the aerosol extinction coefficients derived here and coincident data (from SAGE III, OMPS and OSIRIS). It would be more useful to make this statement quantitative, or at least indicate what criteria were used to conclude that the agreement is "good."

The statement about particle size information would also benefit from clarification: A short summary of the particle size distribution assumptions made and the relative quality of the various retrieved properties would help a reader who is reading the abstract before deciding whether to proceed further.

Line 47: "purposes of the SPS is" – should be "are"

Figure 2: The spectral response functions shown in parts (a) and (b) appear to be fairly Gaussian – it would be useful to include an estimate of the FWHM (or some other indication of their width). The caption mentions that the FWHM defines the spectral resolution, but if those values are stated anywhere, I missed them.

Line 93: This paragraph should conclude with a reference to the later sections in which the non-ideal response of the ALI significantly affects the analysis (to motivate the detailed discussion of this feature, and also to guide a reader who is particularly interested in this aspect of the work).

Lines 101-102: "external source each pixel measured" should be "external source for each pixel measured."

Lines 103-104: Here and elsewhere, the set of images used to characterize the dark behavior is called "large" – how large? And how was the most appropriate large number chosen?

Line 111: "Bad pixels" are mentioned in this paragraph (and defined in the next). Later (in Figure 4), some visual evidence is provided that the number of "bad pixels" is a small fraction of the total image, but how small is it (as a fraction of the total number of pixels)?

Lines 138-139: "Out-of-field" stray light is a serious problem for many limb scattering sensors (usually more serious than "internal" stray light). Is that true for ALI as well? The text says that this is "mitigated with careful baffling" – can you say more about how well this mitigation works, and how its effectiveness was assessed? (If some of these instrument-related questions are addressed more completely elsewhere, then adding a specific reference would be helpful – and I particularly apologize for not reviewing Letros et al., 2024, due to lack of access!)

Table 1: I recommend including the estimated albedo for each scan as another column of data. That would make it immediately clear why Scan 3 includes more highly polarized observations than the others (lack of underlying clouds, which suppresses the amount of multiple scattering present).

Line 155: This paragraph says that the scans presented represent "reasonably nominal conditions" for ALI. Besides the gondola being "relatively stable," how are "reasonably nominal conditions" defined? I assume that the statement about stability refers to the attitude of the sensor – how is this assessed? And do variations in pointing during the integration time of the measurements contribute significantly to the effective altitude resolution of the measurement (or affect measurement quality in other ways)?

Lines 184-185: The definition of the D matrix should be clarified. Would it be fair to call this a Tikhonov regularization term? Stating that it "restricts elements of the optimization from changing too much" should be reworded - I would describe it as being meant to retard the change in the state vector from one iteration to the next in the early phases of the retrieval. (This is reasonable in a Rodgers-type optimal estimation scheme for multiple reasons: An a-priori profile that differs greatly from the true profile may cause the initial retrieval step to move much too far, or in the wrong direction, for example.)

Line 208: "… as discussed later" – this should include a reference to the section where the material appears.

Figure 6: The text provided in Section 4.1 suggests that the albedo estimation process used in this work is new (or at least significantly modified, relative to earlier approaches). But many of the details of the algorithm used are described in the figure caption. This makes the figure caption too long, and fills it with information that isn't directly related to the figure, without providing enough detail about the method itself. I recommend moving most of this caption into the regular text of the paper, and perhaps expanding it (or adding references) to clarify how it works.

Line 226: "… atmosphere of known true state albedo of 0.6." What is the "true state albedo" – the actual albedo of the underlying surface? An effective reflectivity (combining the influences of the surface, clouds, aerosols, etc.)? Or something else? This is related to the following note.

Line 228: "… 0.654 was found, which we consider to be a reasonable estimation of the 0.6 true state." My initial reaction is that this agreement (worse than 0.05, or nearly 10% relative error) is not especially good… but maybe that's unfair. Does the simulated data used in this example include noise, biases, etc. that are meant to mimic the performance of ALI? Does the stated value (0.654) represent an "effective reflectivity" that cannot be expected to perfectly match the "true" value (0.6), for the reason noted in the line 226 note (or for some other reason)? And maybe most importantly, how much does an error in retrieved albedo matter for the retrieval of the aerosol properties? (As noted elsewhere in the text, tangent height normalization often significantly reduces the sensitivity of limb scattering retrievals to uncertainty about the brightness of the scene.)

Lines 237-238: This sentence contains a particularly confusing reference to cloud effects: "… relative changes in polarized light can be used as a metric to determine if limb-scattered signal was influenced by cloud or not." Are we talking about the

"influence" of clouds in the underlying scene here? That's my interpretation, but maybe the statement also applies to line-of-sight clouds? This should be clarified.

Line 268: The reference "Bass and et al, 2010" appears here. That appears to be the only reference that uses "et al." rather than listing the full author list – why? And calling it "and et al." seems redundant.

Line 280: "standard divinations" should be "deviations." And it would be helpful to explain what "prototyping the algorithm" means. I assume this involved experimenting with various settings – you settled on these particular settings for some reason, but how much did the particular selections that you made matter, in the end?

Line 288: "… the θ state is reasonably insensitive" – this should be quantified.

Line 310: Here (and in the caption of Figure 7), the text refers to a "stark" change in DoP behavior (for cloud identification). The figure provides some visual demonstrations that these can appear very obviously in the profile, but how "stark" must a change be to trigger identification of a feature as a cloud?

Line 313: "… known form the true state" – should be "from."

Lines 375-376: "… variance of the median radius is … selected to be 0.01." and "The scalar width has this variance set to 0.0001." Shouldn't these values have units? I may be confused about what these definitions mean… but again, how much do they matter in the observed behavior of the retrievals?

Table 3: How were these listed number density variance values selected, and how do they affect the retrieval? As mentioned earlier, I'm particularly curious how the solution might be affected when the a-priori number density profile differs from the true atmosphere by a factor of 10 or more.

Lines 405-406: "However, of note this a-priori N profile is not constructed from any specific knowledge of the aerosol to be retrieved." What does this mean? I guess the a-priori profile is some kind of climatology, or …?

Figure 10: This caption ends with the statement that "This small error primarily results from the horizontal averaging." For a small error, maybe I shouldn't quibble, but I have to ask: What quantity is being averaged horizontally, and how? Is this averaging discussed elsewhere in the paper, and is it obvious how you determined that this averaging is the cause of the observed error?

Figure 12: The properties of the bimodal distribution are shown in the figure, but do you have a reference for how these particular properties were selected?

Lines 459-460: "This is a similar approach to the standard retrieval approach of OSIRIS and OMPS (Rieger et al., 2019; Taha et al., 2021)." For the latter case, the assumed aerosol size distribution is not log-normal. As stated at the start of the review, I appreciate the experiments with bimodal log-normal distributions to test the approach, but were any non-log-normal experiments also done?

Lines 465: "… yields respectable agreement overall." How is this defined? (With the extinction plots presented on a logarithmic scale, it's difficult to read the percentage error well.)

Line 503: "… atmospheric DoP can be well retrieved" – this should be quantified.

Line 515: "… but we this is a point" – should be "but this is a point."

Line 519: "… extinction in very good agreement" – this should be quantified.

Line 521: "… showed an overestimation of aerosol extinction" – this should be quantified.

Line 503: Is listing data as "available upon request" adequate? I understand that public release of the full set of ALI measurements may not be possible. But inclusion of the data used and illustrated in this study (as a public "supplement" file) seems like a reasonable expectation, which I thought had become a fairly standard practice (for data that is not "officially released" and archived elsewhere).