Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC1
https://doi.org/10.5194/amt-2022-289-RC1, 2022

Comment on amt-2022-289

Anonymous Referee #1

Referee comment on "The CALIPSO version 4.5 stratospheric aerosol subtyping algorithm" by Jason L. Tackett et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-289-RC1, 2022

Review of The CALIPSO version 4.5 stratospheric aerosol subtyping algorithm by Tackett et al.

This description of the changes to the CALIPSO version 4.5 stratospheric aerosol subtyping algorithm is well written and easy to follow with an appropriate amount of detail, and the algorithm changes it describes appear to be logical and useful upgrades to the algorithm. I especially appreciated that almost every time I had a question while reading, the answer came within half a page of my question, which I think is a sign of good flow. The manuscript is particularly strong on explaining the logic of the changes to the classification algorithm. I have only two significant critiques.

First, there isn't much analysis or discussion nor validation of the new lidar ratio, which, along with changes to the classification methodology, was also included as one of the algorithm changes highlighted in this manuscript. I realize that it is nearly impossible to validate the lidar ratios selected for the CALIPSO algorithm, since any available independent information is probably already included in the decision on what range of values to set it at. However, I think some more discussion of the impact and of the sensitivity would be appropriate.

Second, I think there should probably be some emphasis on the idea that the errors in the estimated particulate depolarization ratio are dominated by systematic rather than random errors, and possibly some analysis or demonstration of why estimated particulate depolarization ratio is nevertheless necessary for classification or better than volume depolarization, which is more likely to be strongly dominated by random rather than systematic errors.

Specific comments on content

L58 mentions the algorithmic motivation for getting the right aerosol type is the need to correct the overlying attenuation. How sensitive are the retrievals lower in the atmosphere to having the correct type in the stratosphere? That is, I realize there is a one-to-one relationship between error in layer lidar ratio and the error in the AOT for the same layer, but can the authors describe how much the uncertainty for lower layers is increased given the distribution of stratosphere layer optical thickness? This would address not just correct typing, but also how much impact changing the volcanic ash lidar ratio would be expected to have.

L179-180 includes the attenuated molecular backscatter, "calculated from the MERRA-2 model". How can the particulate attenuation of the molecular bacskcatter be caculated accurately given there is no measurement of the attenuation? Or is the particulate attenuation left out? I hope this will be made more explicit, including some comment about what errors or uncertainties are likely. It seems logical that if the particulate attenuation is not accounted for, then R' will have a systematic low bias (compared to true aerosol backscatter ratio), and estimated particulate depolarization would have a high bias. Is that correct?

L179-180 A more minor point here: the equation and text suggest that the ratio of attenuated backscatter quantities is taken before doing the average. In general, I would think it was better to take the average of each noisy signal and then the ratio, to limit noise magnification. Perhaps this is irrelevant because the denominator is modeled on a coarser scale than the measurements anyway?

Figure 3. I'm curious to see a panel which shows a representative selection of known PSA, which would help illustrate why the PSA selection must be done before thresholding on depolarization.

L274 "based on elevated depolarization". A quick look at the Prata et al. reference that's included in the table suggests that there's supporting information from other sensors. If so, consider including some discussion of supporting evidence in the text here.

L274 "based on elevated depolarization". Given an a priori choice between only sulfate and ash?

Section 4.1 It's very interesting that the color ratio is not used anymore, and this seems to work well. In the analysis it's suggested that there are fewer misclassifications of known cases without the color ratio than with it. However, in Section 4.1 "Color ratio test for smoke removed", most of the discussion is about why the depolarization ratio is sufficient and there is not much discussion about the color ratio. I hope the authors will consider adding more to illustrate why it is advantageous to take it out. For instance, expected uncertaintly levels could be discussed and the distributions of color ratios and percentage of overlap in the color ratio dimension should also be discussed.

Section 4. I believe there's another signature of smoke in CALIPSO data related to the differential attenuation between 532 and 1064 nm (i.e. that the color ratio changes between the top and bottom of the layer). Is this present for stratospheric as well as tropospheric smoke layers, and is there any thought of using this in the automated classification algorithm(s)?

L365-366 "The lowest quartile... was selected so that the average relative uncertainty in estimated particulate depolarization is less than 250%" (and following text). While I'm not suprised by very large uncertainties in estimated particulate depolarization ratio for small signal levels, it's not immediately obvious how an uncertainty of more than 100% is particularly useful, and whether it should make any practical difference if the uncertainty is 200% or 400%, especially given that sulfate has essentially zero depolarization and therefore always has a large relative uncertainty. However, the uncertainty in the estimated particulate depolarization is strongly dominated by systematic, rather than random, uncertainty, particularly for small values of the attenuated scattering ratio, due to the mathematical form of Eqn 2 (see, e.g. Burton et al. 2015) which is a strong and valid reason for wanting to cut off this calculation at values before the systematic error blows up. I think the discussion would be clearer and more informative if the discussion explicitly refers to this systematic uncertainty from Eqn 2 and perhaps includes a figure showing how the resulting uncertainty depends on the attenuated backscatter.

On a similar topic, what are the uncertainty and spread in the volume depolarization ratios? Has there been any investigation into whether it's feasible to do stratospheric aerosol classification using the volume depolarization ratio instead of the estimated particulate depolarization ratio? It would be interesting to see a figure like Figure 6 with joint distributions, but with volume depolarization instead of estimated particulate depolarization.

L409-410 "because knowledge of 1064 nm lidar ratios is not as broad in the literature". But the new 532 nm lidar ratios were arrived at via constrained retrievals, if I followed that correctly. Isn't that technique just as valid at 1064 nm, regardless of other existing literature?

L430 "increased variability in the estimated particulate depolarization ratios for layers [between] the old and new ... thresholds". It's definitely plausible that there is more uncertainty and more variability in these layers. I would argue further, though, that the magnification of the uncertainty and error in the scattering ratio swamps the magnification of the random error in (volume) depolarization measurements (See Burton et al. 2015 Table 2). And if my speculation earlier (see comment pertaining to line 179-180) has merit, then the systematic error will tend to bias the estimated particulate depolarization to higher values. That might account for some misclassification of sulfate as smoke.

L511 please include a reference and/or more explanation for negative brightness temperature differences being a signature of ash.

L618-619 "This reflected sunlight enhances noise throughout the profile overhead, thereby increasing the variability of depolarization ratio measurements". It's clear that noise would increase the variability of the depolarization measurements, including volume depolarization, but why does it produce an asymmetric tail with more large values? Is this because the negative branch is cut off (I mean if negative depolarizations were reported and graphed would there be a symmetric tail on that side?) or is it related to the propagation of error from the volume depolarization to the estimated particulate depolarization? By which I mean, if the volume depolarization distributions are plotted, are they wider during the day but still symmetric? If so, then, considering that noise also increases the variability in R', perhaps the issue is again the R' errors propagating non-linearly through Eqn 2, so that a symmetric distribution of error in R' (plus a symmetric distribution of volume depolarization) becomes an asymmetric distribution of particulate depolarization.

L621 I believe this is a bit of a misquote. Burton et al. 2015 included an overview of other lidar observations, some showing low and some showing high depolarization of tropospheric smoke. Some of the high depolarization ones are "aged", but for logistical reasons, lidar observations of fresh smoke is rare, and of aged smoke is relatively common, so both the low and high subsets are dominated by aged smoke. Burton et al. doesn't give any mechanism for aging affecting depolarization, except to quote Martins et al. 2018 who do. The depolarization of smoke in the troposphere might not be so relevant to depolarization of smoke in the stratosphere anyway. However, note that Haarig et al. 2018 (already quoted in this manuscript) also proposes a hypothesis that's related to aging about why stratospheric smoke tends to be depolarizing while tropospheric tends not to be. And for completeness, there's also the ice hypothesis of Kablick et al. 2018 (Kablick III, et al.: The Great Slave Lake PyroCb of 5 August 2014: Observations, Simulations, Comparisons With Regular Convection, and Impact on UTLS Water Vapor, Journal of Geophysical Research: Atmospheres, 123, 12,332-312,352, 10.1029/2018jd028965, 2018.). Sicard et al. 2019 also have relevant discussion (Sicard, et al.: Ground/space, passive/active remote sensing observations coupled with particle dispersion modelling to understand the inter-continental transport of wildfire smoke plumes, Remote Sens Environ, 232, 111294, https://doi.org/10.1016/j.rse.2019.111294, 2019.)

L656 "The aerosol subtypes of these false layer detections" I'm confused whether these percentages refer to the whole population of false layer detections, or only to the subgroup that remains after filtering for the SAA and low CAD scores. If they include the low CAD group, then is it likely that the dominance of high-depolarization aerosol types might reflect mistyped cloud? Does the percentage classified as ash remain so remarkably high after the low CAD cases are taken out? If so, why should these non-layers be dominated by large depolarizations?

Comments on presentation details

L23-24 "more likely". More likely than what? More likely than the previous algorithm, or more likely to be correctly classified than misclassified?

L112 why "primary"? Are there more resolutions than these three?

L145 I don't know what the "1064 nm baseline shape" means. I understand it is not important to the topic to your paper but since it's mentioned, it would be good to have enough information to understand what it means.

L149-152 Here in the discussion about counting layers, I wondered about whether changing layer boundaries between versions would impact the analysis. Later I can see that it really doesn't, but an explanation in this paragraph about why lack of perfect layer matching doesn't impact the analysis would be helpful.

L274 "their composition was determined". Add references.

Figure 5 flowchart. In the decision about the attenuated backscatter on the top row, third diamond from the left, I believe < should be > here.

L362 "confidently classified" is confusing since the reason for the cutoff is that they can't be confidently classified. Did "confidently classified" refer to the cloud-aerosol discrimination? Can this be made clearer?

L368 "those layers" I'm having trouble following. Which layers? Does this refer to daytime/80km layers that have attenuated backscatter between the old and new cutoffs? If they can't be confidently classified, should they be left in the unclassified category?

L369 "these classifications". Again, which ones? The 80 km layers? Or all the layers between the old and new thresholds?

L372 "increase of 50% in the relative uncertainty". This is ambiguous. It could mean a relative increase of 50% or an absolute increase of 50 percentage points. I think the latter is what's intended. If so, consider "50 percentage points".

Figure 15d. Why is the time series cut off while the sulfate signature is still high?

Figure 16. It's somewhat confusing that this shows volume depolarization ratio when the

text primarily discusses estimated particulate depolarization ratio.

L673. "threshold ... to discriminate between volcanic ash and sulfate". I think this is a typo, as these two types do not share a boundary.

L693. "smoke transported". Consider "any smoke transported" or "if smoke is transported"

L708. Add reference for the split window technique.