Review of

Using OMPS-LP color ratio to extract stratospheric aerosol particle size and concentration with application to volcanic eruptions

By Yi Wang et al.

Summary

This paper is both poorly conceived from a scientific perspective and poorly executed in the mechanics of producing a paper. While I have made numerous potential corrections for the manuscript below, the scale of these changes cannot be reasonably be contained within this review process. As a result, I cannot recommend any course except a complete rejection of this paper.

- The writing is poor and at times impossible to follow.
- I don't normally worry too much about references, but I noted several errors in passing so I checked more thoroughly than I've ever checked them before. I found numerous bad attributions and misinterpretations of other work that always seemed to support the views of the authors. It seems like the authors made little effort to verify that references corresponded to what is said.
- The authors own misunderstanding of what they are doing is reflected in the use of SAGE III 'validation' which is not validation in any sense of the word. I was dumbfounded when I read it.
- The authors never address OMPS data quality for which there is ample reason to be concerned particularly in the enhanced periods they address in section 4.
- Though I was not a reviewer of the earlier version, it appears to have very limited changes relative to concerns raised earlier in the review process.

Major points of concern

The technique they use for inferring mode radius and number density are virtually identical to the methodology developed by Glenn Yue in 1983 (except perhaps the error discussion). The introduction of the technique must acknowledge this heritage up front and then may expand on other unique extensions to the original method (158 and onwards). The discussion includes a description of the algorithm used internally to the retrieval algorithm to produce the aerosol product (58-59). I am left with the feeling that this is intended to replace that methodology though there is no further discussion of this. Obviously, if this is intended then the first step would be to show how it impacts the OMPS data product. If this methodology is not intended to replace the current internal approach, then the language in this section should be cleaned up including simply not discussing how aerosol is derived internally. The discussion of how the OMPS aerosol product is produced in mentioned in multiple different places (e.g., 59-62, 75-79, 108-110) and if it is to be discussed at all (and I am not sure I see the relevance to the topic of this paper) then it should be combined into a single coherent discussion.

The use of SAGE III observations as 'validation' is completely without merit and must be removed in its entirety (232-269). Given that the same retrieval algorithm is applied to both data sets (OMPS and SAGE) and the similarity of the measurements and relative simplicity of the retrieval methodology, the only 'validation' possible is the comparison of aerosol extinction coefficient, their ratios, and their comparable uncertainties and thus a traditional instrument comparison that is out of the scope for this paper. Beyond the agreement between the actual measurements (extinction coefficient) results for number density and mode radius are preordained. This is not validation of the retrieval approach in any way. This entire section should be removed since it has no value and is prone to deceiving readers that it is real validation.

The authors make a good point that the community is likely approaching a period without solar occultation data to support limb-scatter measurements of aerosol extinction coefficient in the stratosphere (48-56). They also mention a number of factors that make the limb measurements more difficult than occultation. However, they make no mention of the outcome of these limitations particularly during periods of enhanced aerosol which are the focus of the second half of this paper and, of course, are the most interesting scientifically. This is especially concerning since it is well known (e.g., Bourassa et al http://10.1029/2022GL101978) that OMPS aerosol extinction data does not agree well with occultation (SAGE III/ISS) data or comparable limb measurements by OSIRIS following volcanic and smoke events (OD out by a factor 2). This is almost certainly related to limitations of OMPS instrument and algorithmic performance. The lack of transparency on OMPS data quality is deeply concerning as non-experts use observations with the expectation that they are as described by the mission team.

On a similar note, the comparisons early in the paper (OPC and SAGE) show comparisons for fairly benign conditions yet the application focus of the paper is on the Raikoke and HTHH eruptions. These are conditions where assumptions about aerosol are the most likely to deviate from the 'normal' values for width and composition that are made as a part of this retrieval process. While extinction coefficient between OMPS, OSIRIS and SAGE are broadly decent during benign periods, it has been shown by Bourassa et al. (http://10.1029/2022GL101978) that the agreement in extinction with more direct measurements like those of SAGE are the worst during perturbed conditions like these. This is most likely due to breakdowns in the assumptions in the retrieval of the extinction coefficient products. If the authors really want to show how their retrieved number density and mode radius perform for these periods, they must also validate their results for these more challenging periods. Otherwise, the results are highly questionable and possibly prone to misinterpretation by readers.

Other points of significant concern

It is 'aerosol extinction coefficient' not simply 'aerosol extinction' which is the complement of transmission. Rather than referring to aerosol 'size' the authors should use the correct terminology and use 'mode radius.'

References on alternate aerosol compositions of ash, smoke and mineral dust includes two that reference ash (Vernier et al. 2016; Muser et al., 2020) and two that reference meteoritic material (James et al.2023; Schneider et al. 2021) which is not even in the list. This should be corrected, and they should include organics as well. (26-28)

The following sentence (33-35) is wrong at several points: *The solar occultation technique directly measures the aerosol optical depth at multiple wavelengths, usually 300 – 1000 nm, and thus can estimate the path average aerosol concentration and size (Thomason and Taha, 2003).* First, solar occultation's primary aerosol product is aerosol extinction coefficient not optical depth. The technique can be (and has been done) at 1 or more wavelengths. They do not in anyway directly measure aerosol size or number density. Those parameters have been frequently inferred from the measurements (in possibly dozens of papers) but that is a separate activity with its own strengths and limitations.

If the authors want to make the point that solar occultation produces much less data than limb scatter that is welcomed. However, they should explain how they both collect data and relative strengths of these methods instead of attributing it to 'special geometry' (46-47). For that matter a brief overview of how limb scatter infers aerosol extinction would be appropriate as a part of a compare-and-contrast discussion.

Thomason and Vernier (2013) applies to identifying clouds in the tropical UTLS using SAGE II data and does not apply to size distribution retrievals (64). It is possible that the authors meant to reference Thomason et al. (2008) which used two channel retrievals to bracket the uncertainties in bulk property retrievals like surface area density but not number density or mode radius.

Schoeberl et al. (2021) and Kovilakam et al (2022) (65-66) (the latter of which needs its citation to be updated), apply to cloud identification and not two channel size and number density inferences. Since the primary author here is also a coauthor of the Schoeberl paper, one would presume that they would be aware of this.

The discussion of aerosol extinction calculations beginning at line 88 and thereafter is torturous. Equation 1 is an extreme shortcut that is totally inadequate. They should begin with the full Mie integral equation of computing extinction with a generic size distribution, note that they will hence forth assume a single mode log-normal size distribution and place that in the equation, from there, it is mathematically obvious that aerosol extinction coefficient ratio (color ratio) is insensitive to the total number density. This is true for any aerosol size distribution but really obvious once the SMLN is substituted into the Mie equation. It should be made clear that the color ratio IS sensitive to the relative distribution of aerosol across the radius domain, that being the reason that color ratio varies as the aerosol size distribution varies. That said, much of the following section down to about line 113 can be removed including Figure 1b (which is really pointless in any case). Figure 1a is important and can be folded into the discussion I suggest above though I would probably rotate the figure so that the independent axis (x) is color ratio and radius is the dependent variable since this is how the retrieval algorithm works. It should also be pointed out that the slope of the two curves shows why the wavelength range of measurements is so important since small changes in the 745/869 ratio suggest much larger changes in mode radius than similar changes in the 510/869 ratio. Figure 1b really only shows the obvious and can be deleted. It is obvious to state that you can't derive 2 variables from 1 measurement as stated in the current section.

The measurements of interest for this study are listed as 510, 745 and 869 nm. Only for 869 nm measurements is a justification listed for selecting it (low interference from other species). Clearly, the contribution of ozone and molecules at 510 and 745 nm are considerably higher than at 869 nm so I don't understand why 869 nm is called out in particular (92) and why the others do not merit discussion. In any case, the authors should provide a list of ALL aerosol measurements made by OMPS and justify their selections for use in this study. Most papers I have seen using OMPS aerosol data use the 997 nm measurement and refer to it as the most robust OMPS aerosol measurement (including ones referenced herein) yet in this paper, this measurement is not even referred to. Has understanding of the OMPS measurements changed or is there some other reason for this omission?

The cloud discussion should be improved (115-122). I'd guess that with a cloud-top measurement that further cloud identification is because some clouds leak past the filter. There is no justification given for using a color ratio of 1.1 to filter clouds (unless it is from the reference in the previous sentence). I assume that the only relevant color ratio here is the 510/869. It is not at all obvious in the figure that 1.1 is the appropriate value. It is clear that the ratio decreases in the vicinity of the tropopause but it is impossible to use the figure to justify this. That values much smaller than 1 appear suggests data quality issues rather as they are impossible physically. I assume your conclusion is that optically thin clouds are removed though this is not demonstrated.

I think it would be better to introduce the uncertainties in color ratio and width before introducing equation 3 (147-149) although this is basically just computing the standard deviation so doesn't even require an equation.

Equation 4 makes no sense to me (line 152). The implication is that the relative uncertainty in aerosol extinction coefficient is one-to-one related to uncertainty in radiance. Since aerosol is not the only radiatively active component of the atmosphere, and at the shorter of these channels it is most often not even the largest component in the measurement, I am confused by this, and I suspect that the product uncertainty must be larger than the base radiance measurement uncertainty. It is certainly that way for every other measurement I am familiar with.

It seems like the sentence at 160-161 suggests larger uncertainties in color ratio and width at some altitudes then used in Figure 4. This would suggest larger uncertainties in mode radius (and number density). How much bigger? That is followed at 161-162 with a summary of uncertainties for I assume the nominal error levels (please clarify this). The order of these sentences should be reversed with the sentence at 164-166 between them. Please discuss how much bigger the uncertainties can get with these new errors (where do they come from? Altitude dependence?). I would delete the summary sentence (162-164) as the rest of information is a few sentences above.

The uncertainty in width is given as +/- 0.2 and from Rieger et al (2018)'s Figure 6 (154-155). This plot has two modes (fine and coarse). In any case, the range shown in Figure 6 might be +/- 0.2 from the mean at the best altitudes but is often much larger. In addition, the nominal 1.6 which is used in this paper ranges from the lowest value in the range in Figure 6a (middle) at 10 km to the upper bound between 20 and 25 km (where the range is the smallest), a factor not mentioned by the authors. I do not see how the authors can use this figure to define the range as 0.2 nor does it support the use of 1.6 as a universal width.

If Figure 4 (158) is produced only using data from Figure 3, then how can the plot span the same range in the variable space since the grid points are produced using a finite box of +/- 0.2 in width and 21% in color ratio. At the edges, this box runs off the regions covered by Figure 3 so computing a standard deviation isn't possible for widths less than 1.3 or greater than 1.6 or for color ratios less than about 1.4 or greater than about 6.5. If that is done effectively the statistic changes character at the edges of the plot. So, this looks wrong to me.

The propagation of uncertainty in mode radius to number density is extremely simplified (164-166). One, the uncertainty in the absolute extinction coefficient (not just the radius) must also play a role in the resulting number density uncertainty. In addition, while at the largest aerosol sizes, extinction coefficient scales well with radius squared this is not the case at smaller aerosol sizes. Since the size distributions span a large range of radii and extinction efficiencies. The impact of radius uncertainty on the inference of number density can be done in a straightforward way or simply simulated with the authors Mie code with minimal effort.

I have no idea what the authors are trying to say in the paragraph at 172-176.

I do not see the point of section 2.3 (177-202). It looks like a discussion of the sensitivity of the aerosol product itself as opposed to the mode radius and number density inferences that are otherwise the focus of this paper. I don't understand the relevance. I also don't understand why perturbing the α and β parameters by 10% is relevant. Do the authors have some reason to believe that such perturbations are reasonable? Overall, it seems like they are randomly perturbing the gamma distribution parameters and then treating the random results as if they mean something. The whole section MIGHT be worthwhile if the authors compared the differences in phase function (at relevant angles) between the inferred size distributions and the one used in the retrieval of the data products and following those differences to differences in the primary data products themselves but moved to an appendix. Otherwise, this section seems to be a cousin of the subject in this paper.

Equation 8 (216) looks like the error in color ratio that I was expecting to see where Equation 4 is. Why wasn't this used before. I assume, though the authors do not state this, that the uncertainties in the number density and mode radius are now computed this, apparently different, measurement uncertainty. This discussion is a little unclear.

The base product for the balloon data is number density greater than a set of radii. I believe that these data sets use a bimodal aerosol size distribution in the production of the base product and may be available in conjunction with the other data. The larger mode is not always significant, nonetheless, I wonder how the authors come up with a single width and mode radius from this data set. Is this a regular product (a single mode fit by the Wyoming group)? Is this only the data for the fine mode? Are the authors doing their own fit? It must be made clear how they are using this data.

In Figures 7 and 8, it would be very helpful to have the relative differences shown for at least plots 7c, 7g, 8c, and 8g. Particularly for the multidecade log scale it is hard to get a feeling how close the two products are.

Like one of my concerned regarding the use of the SAGE III data above, I note that both profiles shown in this discussion (210) are for very benign conditions with no recent volcanic or smoke events. This is despite the key application of this new approach being to anomalous events in the stratosphere. While the Wyoming data (taken there) is unlikely to have seen the HTHH plume, it must assuredly have observations of the Raikoke event. Given the likelihood of substantial size distribution changes between benign and enhanced stratospheric aerosol, it is an absolute necessity that the authors provide this comparison.

Figure 11a is very similar to Figure 6 in Thomason and Vernier (2013) which the authors have previously referenced. Both show a tail from the aerosol 'core' to optically thin clouds however, in the Thomason and Vernier paper, the entire tail is interpreted as being influenced by cloud (this also shows up in GloSSAC publications) whereas only a core is attributed to clouds in this paper. How confident are the authors that their very simple cloud identification method is functioning?

Most of section 4 is OK though, though this section cannot be fully reviewed due to the shortcomings earlier in the paper.

Section 5 (conclusions) repeats much of the flawed thinking earlier in the paper and would have to be completely revisited if a new version is attempted.

Minor points

There are a number of undefined variables such as E (line 75, defined at 88), W (149 which I think is width)

'impact on climate' not 'impact of climate' (18)

It would be more correct to say 'aerosol properties vary over time due to' than directly associating the changes to number density as if it was the only relevant parameter (20-21)

Sulfur not sulphur (26)

Lofting of sulfuric acid aerosol due to absorption of infrared radiation also occurs. It is also likely more complicated than creating 'dark' (more properly absorbing) aerosol, as the nuclei can dissolve in the sulfuric acid aerosol as well as remain as an inclusion and may not be included within that aerosol at all and self-loft without sulfuric acid components. (28)

Solar occultation is not the first remote sensing of aerosol in the stratosphere. That honor belongs to search light studies by Elterman (Applied Optics, 11, 1769, 1966) and early lidar work (not the best reference but they are around: https://doi.org/10.1175/1520-0450(1971)010<0443:ASGLR>2.0.CO;2) (33)

SAM II not SAM-II (37)

I am unaware of any stratospheric aerosol measurements using the Moon as the target (38).

I think the authors are using the only the Mie component of the SASKTRAN radiative transfer model rather than the entire code and that is not fixed to any particular size distribution width. (81-86)

The exponent in equation 2 is wrong (135) missing both the squared and a 2 in the divisor. I presume this is a typo.

Figure 10 (169) should not be referenced before Figures 5 through 9. Probably this discussion should be labeled 'as shown below' or entirely shifted to the section on Raikoke and HTHH.

Using Ward et al (208) as the reference to the Wyoming OPC is kind weak (not wrong) since it is about PSCs and doesn't have Deshler or Kalnajs as first author.

Figure A1 (179) should not be referenced in the main body of the text.

A period is missing at the end of equation 6. (185)