I want to thank Dr. Pohl and her coauthors for the work they have done in writing this paper. Further, I want to thank them for the work that went into this revision. Finally, I want to thank them for putting up with my myriad questions and comments.

Overview

This is a challenging paper to review because of its size and the complexity and detail of the analysis. There are multiple variables at play throughout and correctly interpreting the results requires considering the data from multiple facets. With that in mind I readily acknowledge that I may have miss interpreted something, which has led to my confusion. However, I must also admit that I struggle to see the utility of this method for the reasons below.

Major Issues

This methodology is highly sensitive to the assumed, a priori, number density profile as demonstrated in Fig. 2. Here, the authors showed that changing the a priori N profile (by factors of 0.5 and 2) changes r_g by $\approx \pm 30\%$, σ_g by $\approx \pm 6\%$ (no big deal), r_e by $\approx \pm$15-20%, and extinction by -45% to +140% (it is interesting to note that this scaled nearly linearly). Granted, all bodes well when the a priori N matches current conditions (as the authors demonstrated in Fig. 5 and elsewhere). However, under volcanically active conditions it is entirely reasonable that the a priori is more than a factor of 2 different (looking at the Wyoming OPC record I see changes in excess of a factor of 10 after eruptions within the SCIAMACHY time period). Taken to an extreme, how would this method perform after Pinatubo or Hunga Tonga?

While the authors evaluated the influence of an incorrect a priori (via Fig. 2) there remains 1 glaring shortcoming of the method: the "real" N profile is unknown therefore we do not know how much uncertainty this introduces to the retrieval and we cannot quantify the uncertainty of the inferred PSD values and the derived extinction coefficients. What we do know is that this uncertainty *can* be substantial. If the authors were to limit their analysis to conditions when N is stable then they could make reasonable guesses for their a priori (that's basically what they do here, using the OPC record). However, that is not interesting. The interesting bits are in the post-eruption atmosphere when the N profile and PSD parameters are **most** dynamic! In short, we know that the PSD parameters and extinction coefficients as derived from SCIAMACHY data are wrong. . . but we have no gauge for how wrong they are. Unfortunately it's not *just* the quantitative results that are suspect but we must also suspect and qualitative interpretation of the data as well.

I really like this paper so this leaves me with a dilemma. I am left questioning how I would use this data and what is the ultimate purpose of this paper (i.e., what does the community now know that we did not before). I think what we now is this: PSD parameters can be inferred from SCIAMACHY data and these inferred parameters have modest sensitivity to the a priori N. The calculated extinction is much more sensitive. However, given the extreme range of N in the Wyoming OPC record after major eruptions within your time period I have to conclude that this methodology is useful only during stable/background conditions and not reliable in the aftermath of volcanic eruptions. This limitation is systemic throughout the paper and is inherent within the methodology itself and I currently see no path to salvaging it. It is for this reason that I cannot recommend this paper for publication.

I recognize that my view may be in the minority and, should the editor decide to allow publication, then I fully support him in this decision.

Specific and Minor Comments

- **page 7, line 197:** "unambiguity" should be "ambiguity"?
- **page 9, lines 268–269:** "the retrieved PSD parameters and the assumed number density are used to calculate the effective radius (Eq. (5)) and the extinction coefficient (Eq. (6)) of the aerosol particles" Earlier in the manuscript you stated that N can be fixed through space/time because N plays a minor role in the retrieval process (fair enough). However, here you see how N plays a crucial role in calculating some derived parameters (especially extinction). Perhaps a statement regarding this dependence is appropriate here.
- **page 10, line 283:** "classes" should be "bins"?
- **page 17, lines 475–479:** "A quantitative error estimation of both assumptions, the Lambertian surface and the a priori N profile, is therefore not possible for real retrievals. We can only point to these sources of uncertainty." This is an accurate statement and it is highly unfortunate. In my view, this is the dominant shortcoming of this method: you know the numbers are wrong but you don't know by how much. This uncertainty will be more pronounced immediately after major events (N can change by a factor of 10 or more, which is FAR more than the "doubling" you modeled). It may be necessary to explicitly tell the reader of this shortcoming post-eruptions.
- **page 18, lines 483–484:** In what way is this "striking"? Fig. 2 showed that underestimating number density (yellow line) results in over estimation of particle size and under estimation of distribution width. This results in an over estimation of approx 100% in the extinction coefficient and an over estimation of approx 20% for r_e. Undoubtedly, the a priori N value in your model is too low after these eruptions, which puts you squarely in the situation I just described (i.e., over estimation of r_g, over estimation of extinction, etc.). I don't doubt that extinction increased, I don't doubt that particles became bigger, and I don't doubt that distribution width decreased. However, particles do not always get bigger after eruptions as some of your co-authors have demonstrated (https://acp.copernicus.org/articles/23/9725/2023/). Therefore, this leaves the reader wondering how much of the variabilty shown in Fig. 3 is a by-product of a wrong number density. Given the level of uncertainty in this method I do not believe that you can say that your data unambiguously proves (much less quantified) these changes occurred. Undoubtedly changes are expected, but at this point I think that the most defenceable statement that can be made, based on your product, is that things changed. . . by some amount.
- **page 19, Figure 4 caption:** The OPC record reports 2 modes (fine and coarse mode). Did you use both modes in calculating the extinction coefficient? The coarse mode can have a disproportionate impact on extinction.
- **page 19, Figure 4:** Panels (d) and (i) do not make sense. Why do the red and blue lines cross each other at ≈18 km at not at ≈20 km (i.e., where the 2 a priori N lines cross each other in panels (c) and (h))? All other panels have the red/blue intersection at the same altitude so why are (d) and (i) different?
- **page 19, Figure 4 caption:** What about the light red and light blue colors? Can you define those here so the reader need not search the text for the explanation?

- **page 20, lines 522–524:** "Remarkable are the similar profile shapes from SCIAMACHY and OPC on 7 November 2009 in case of r_g and σ_g by assuming the a priori N based on balloon-borne measurements. This is due to the similarity of the SCIAMACHY-assumed and OPC-measured N profiles." This is no surprise (as the authors state, this is due to the extreme similarity between the current OPC N profile and the climatology, that was based on OPC data). What this tells me, yet again, is that the profile is entirely dependent on the a priori N profile.

- **page 20, lines 526–527:** "good agreement of the extinction coefficient from OPC and SCIAMACHY, regardless of the assumed a priori number density (Fig. 4(d,i))" I disagree on the interpretation of this figure, but I admit that I am struggling to find an interpretation and more information would be helpful. Panel (d) certainly looks promising (SCIAMACHY and OPC are in good agreement) and I would be surprised if it were bad. However, panel (i) is less impressive. While the shapes are in good agreement the OPC extinction is larger by a factor of 2-3 (is this what is meant by good agreement)? Undoubtedly all of this variability is driven by differing number densities. However, it is important for the reader to understand how the OPC extinction coefficients were calculated here: did you only use the first mode or both modes? If you only used the first mode then, especially after an eruption, we can reasonably expect the second/coarse mode to be enhanced and have a disproportionate impact on extinction. Here is the point: currently the difference between SCIAMACHY ext and OPC ext is 200%-300%, if you include the second OPC mode in calculating ext the difference will become larger and this does not qualify as "good agreement".

- **page 20, line 528:** "...the three PSD parameters remain consistent with each other." I do not understand what is meant by "consistent." Can you please clarify?

- **page 21, Figure 5 caption (and corresponding text):** "...relative errors..." This assumes that the OPC is correct, which it is not. You stated on the previous page that disagreement between OPC/SCIAMACHY was driven, in part, to differing sampling volumes and because the OPC only sampled the edge of the aerosol plume while some of the SCIAMACHY profiles were collected within the plume. Should be reworded to "percent difference" (or something comparable) here and throughout the text.

- **page 21, lines 539–540:** "Since SCIAMACHY is not sensitive to stratospheric aerosols with r_g ¡ 0.06 μm (Malinina et al., 2019), corresponding OPC PSDs are excluded from the comparison" This is not correct and these smaller r_g's should be included. They will not contribute much to extinction at 750 nm, so no big deal, but, as written, you are not doing a valid comparison.

- **page 21, lines 547–548:** "...Ext depend only slightly on the choice of the a priori N profile." This is not what Fig. 2 tells me; N is highly important.

- **page 22, lines 566–567:** "...to obtain a sufficient number of collocations." Sufficient for what? You now have 4255 coincident profiles...how many did you have when the maximum distance was only 200 km?

- **page 23, Figure 6:** This behavior is exactly what I expected based on Figure 2.

– **page 24, line 585:** "Discrepancies are smaller in the middle latitudes. . . " Yes, this makes sense because these latitudes are the least impacted by volcanic activity. Therefore, your a priori N profile is more similar to the N profile of these latitudes.

– **page 34, lines 774–775:** "The retrieved median radii and geometric standard deviations should therefore be considered with caution in areas with high aerosol loading." This is true for extinction as well.