We would like to thank the fourth anonymous reviewer (AR4) for providing feedback on this manuscript. Our responses are provided below (red) to AR4's comments (black).

The authors did not discuss what the results of the classification method will, or could be, used for. Large wildfire smoke events may be possible to identify in periods of moderately volcanically elevated sAODs, but OMPS (UVAI, ext coeff) and Calipso (dep ratios, col ratios) already does this. Are there any advantage of using SAGE rather than other platforms?

While other instruments are routinely used to observe wildfire and volcanic activity, it is important to understand the applicability of SAGE data to these identifications. While coincident observations (in both time and space) strengthen the interpretation of each instrument's data, it is not always possible to have multiple instruments observe the same location within a reasonable time frame. Therefore, it is necessary to be able to understand the strengths and limitations of each individual instrument. To this end, we presented and evaluated a method of distinguishing between sulfuric acid aerosol and smoke in the stratosphere that uses the SAGE extinction spectra. The introduction was updated to communicate this.

Can you be sure that it is only wildfire smoke and not something else? Carbonaceous components have been found in volcanic aerosol. Could it not be that one of the eruptions of Raikoke contained soot and organics, or that smoke in the area of Raikoke was entrained in the volcanic cloud?

We apologize, but we do not understand what part of the manuscript this comment refers too. If this is just a general comment, then our response is: We discuss in the paper why some misclassifications may take place, though we see no evidence for soot and/or organics being a significant component of the Raikoke ejecta. The misclassification discussion has been updated in the current manuscript.

Further, I found no difference in depolarization ratio for the assumed wildfire smoke and volcanic aerosol (after Raikoke). Why does it not show up as wildfire smoke in calipso's depolarization ratio? You show information on particle sizes, but not on other particle properties. I think that the depolarization ratio should be shown here since it is a very strong indicator of smoke.

We apologize but we do not understand what the reviewer is referring to. If this is in reference to Figure 14, this figure demonstrates that smoke was in the stratosphere prior to the Raikoke eruption.

Section 5: In the beginning of this section, it reads that the slope was computed via linear regression. How did these regressions look and how well did they fit to the data? From the figures, e.g. Fig4, it looks like there is a large variance in the data. It is difficult for the reader to grasp this without some type of illustration of these regressions. Aren't the widths of these distributions rather important for your classification? The standard deviations of these linear regression models could be used to distinguish between cases where the identification is more or less 100% indicative of one class, and cases where the data points are mixtures of smoke and sulfate. I think that this could be a means of telling whether the rising stratospheric aerosol after Raikoke/Ulawun/Fires are a mixture of smoke and sulfate or only smoke.

It is correct that slope was computed via linear regression, though the residuals of this fit were not part of the calculation and subsequent analysis. There may be some confusion here. Figure 4 used extinction ratios and not slopes. This figure is not part of the analysis, but was a preliminary step we used to observe the general behavior of the SAGE data to see, to a first approximation, whether the data behaved as expected from theory (i.e., based on Fig. 3).

Regarding the "standard deviations of these linear regression models", again, we apologize but we do not see how standard deviation of a linear regression plays a role here.

I think that the figures showing the altitude dependent slopes are really good illustrations to highlight where the different aerosol layers are located. Does this work well when separating background aerosol from low volcanic impact?

As demonstrated by the case-study events this works well when the criteria on lines 266-272 are met, as described in the original manuscript. Ultimately this depends on how much the "low volcanic impact" events change the extinction coefficient.

In the analysis you had to divide data into altitude segments (since extinction coefficients in rising or descending air masses becomes pressure dependent). Would it be possible to normalize the data with pressure to get an altitude independent slope for each class (backgr, volc, smoke)?

Data were broken into altitude ranges because the background aerosol load changes as a function of altitude. Therefore, we had to develop an altitude-based statistical set for each location to determine whether or not a plume was enhanced). This statistical set may be used at different locations within the same latitude band, but the altitude grid must be retained. Broad application (in both altitude and latitude) of a statistical data set is ill advised in this scenario.

Table 3: It looks to me that there are quite some misclassifications even at times and altitudes with large sample sizes. Starting with the Canadian fires 2017, 62% of the data at 14 km altitude were classified as sulfuric acid, and at the highest altitudes (23-25 km) 57-99% are misclassified as sulfuric acid. What would be the source of this sulfuric acid? I don't know of any potential eruptions occurring in the first half of 2017. To me this indicates big issues with the assumptions used for the classification algorithm. The same issue occurs after the Australian fires 2019/2020, but only at the highest altitude shown (25 km). I would like to see how well the algorithm does above 25 km. It is evident in the SAGE 3 iss data that the smoke rose to >30 km, and some dense smoke layers in the v5.10 data lacked data below 27 km indicating too high optical depth in the line of sight to quantify the extinction. So these are not faint layers. It is difficult to interpret the classification results after Raikoke if these issues occur in the periods of known sources.

We do not propose a sulfuric acid source for these time periods. Rather, we discussed potential reasons for misclassification of these very events from lines 334-354 of the original manuscript. This discussion has been expanded in the revised version.

It is true that smoke was detected to high altitudes after the Australian fire. However, we limited the scope of our analysis, as presented here, to 14-25 km for 2 main reasons. 1. we used extinction coefficients from 450-1550 nm in this analysis to calculate spectral slope. The sensitivity of longer-wavelength channels decreases rapidly above 25 km. 2. Considering the case study events as a group, there were fewer interesting enhanced layers in this altitude regime.

After Raikoke all the highest extinction coefficients (Fig 15) were classified as smoke. I find this surprising. Does this mean that a large fraction of the AOD elevation after Raikoke was actually caused by fires?

This is an excellent point and we thank the reviewer for raising this issue. No. Just because a layer is identified as "smoke" does not mean it is composed 100% of smoke, and it does not mean that that layer is majority smoke. This is important and we thank the reviewer for raising this question. We have added additional comments regarding this interpretation and the potential for misclassification to the text.

I wonder to what degree the small difference in refractive index affects the classification. The refractive index for black carbon differs quite from that of H2SO4. However, brown carbon and sulfuric acid has rather similar values in refractive index, indication that it is difficult to separate between the two.

The 2 smoke curves show a range of potential values that are dependent on the composition (or degree of "complete" combustion) of smoke. The actual refractive index for smoke is highly variable as shown be Liu et al. 2015 (now reference in the revised manuscript), and the refractive indices we chose provide a reasonable representation of the BrC RI boundaries in Liu et al. 2015's Fig. 4. As stated above, wildfire burns result in a mixture of BrC and BC being released into the atmosphere and the BC/BrC ratio will be highly variable depending on burn conditions. Further, the composition of BrC determines its spectral properties (i.e., refractive index), which results in a wide range of possible refractive index values (as now discussed in the revised manuscript). Of course, this is all complicated by the lack of in situ measurements of stratospheric smoke. Indeed, it would seem that there is a great measurement and modeling opportunity here that should be seized, but is outside the scope of this manuscript. This figure was updated to show how addition of a small amount (10%) BC significantly changes the overall slope as compared to the pure BrC curve. The consequence of this is that if a smoke plume is composed of 90% BrC and 10% BC and has a nominal mode radius of 125 nm (on the lower end of what is expected for smoke particles) then the resultant slope is easily distinguished from the slope yielded by background sulfuric acid aerosol. This explanation has been added to the manuscript for clarity as has additional language on the use of this figure as a predictive model. This figure presents a general relationship and while we expect this general relationship to be valid, we explicitly state that stratospheric smoke is more complicated than this simple model and that the case study events stand on their own, independent of this figure. Further, the manuscript now contains expanded discussion on the possibility of misclassifications.

Why did you not include a spectral slope for ash (Fig. 3), and in what way may this impact your classification?

Much like smoke, ash has a highly variable composition that results in variable refractive indices. However, we did discuss in the original submission that ash should behave like smoke due, primarily, to it's larger size. As shown in Fig. 3 of the original manuscript, as particle size increases the spectral slope tends to flatten. Therefore, we expect it to have a flatter slope than sulfuric acid aerosol, and will therefore be misclassified as smoke in the current method.

The numbers in Table 3 don't add up. I did not check them all but noticed the issue at

Australia @25 km altitude (0.30 + 0.60)

We are appreciative of the reviewer's keen eye and pointing this out. This was corrected.

L325-328, regarding Fig 7&8: You claim a rapid decrease in the slopes. I see a slope changing value over several kilometers. Isn't this an indication of mixed sources?

The slopes for the plumes indicated in the text changed rapidly over an altitude range of 2.5 km (Fig. 7) to 1 (Fig. 8) km. Similar enhancement in extinction coefficient is observed in panel (c). While the width of the plume is ≈5 km thick, the rate of change from background conditions was high. To the reviewer's point: we see no indication of a mixed source in the Ambae and Ulawun events.

Regarding the figures with calypso curtains, I suggest that you add curtains of the beta-532 signal as it is difficult to understand why there is a yellow feature in Fig 7b (same in Fig 8). The volcanic layers should be visible in beta-532.

We thank the reviewer for this suggestion. CALIOP backscatter products were added to these figures.

L358: You write about an aerosol layer at 19 km altitude. It is actually visible in calypso images, but it is classified/misclassified as clouds. No cirrus should be present so far (7-8 km) above the TP (even above the ExTL).

The reviewer is correct that a layer was identified in the CALIOP VFM ≈5to the north and ≈7to the east of the SAGE overpass location, but it was not visible where SAGE sampled the atmosphere.

Section 2.1: Why was the data limited to 2 km above the TP? Was it to minimize cloud interference? And why not 1 or 3 km? Are there any risk of cloud interference that may disturb the classification?

Correct, this limited the impact of cloud interference. While cloud interference is always a concern near the tropopause, we see no evidence of cloud interference in the data within the selected range.

Section 2.2: The lvl3 sAOD product has strong bias in the extratropics (Kar et al. 2019). Does this have any impact on the comparison with SAGE?

We apologize, but we do not understand the intent of this comment. The L3 CALIOP sAOD product was not used in this analysis.

Caption Figure 3: I think that you should add the word 'normalized' to the ylabel.

We appreciate the suggestion, but this information, in addition to an explanation of how normalization was done, is in the caption.

There is strong gradient in the slope in Fig. 12d at 10-11 km altitude. Could it be clouds interfering? Also, no TP was marked in Fig.7&8. Is the TP height lower than what's shown in the graphs?

No, this is well above the tropopause (8 km), which can be seen in panels (a) and (b) of these figures.

The particle size distribution evolves with time, especially in the first month or two after eruption (or smoke injection). This should lead to increased variance in your data.

We thank the reviewer for this comment, but we do not see anything actionable from this.