1 Response to Reviewer #1

We would like to thank Mike Fromm for providing feedback on this revised manuscript. We also appreciate Dr. Fromm's willingness to engage in fruitful conversations about this manuscript via e-mail and telephone. Our responses are provided below (red) to Dr. Fromm's comments (black).

The manuscript has undergone a substantial revision so some of the comments/responses may not be applicable to the newer version. The revised manuscript presents this analysis from a different perspective where the 4 case-study events stand completely on their own and demonstrate that this method works well on pyroCb events and small/moderate eruptions. The introduction was updated to help the reader distinguish between small/moderate eruptions (like Ambae and Ulawun) and larger eruptions like Raikoke: the primary difference being mass of ejected SO_2 and *not* VEI rating. Further, after demonstrating that the method worked well on the case-study events we present the Raikoke eruption as an example of an event that falls outside the applicable paradigm of this method (i.e., Raikoke released enough SO_2 to create large particles). We removed language that referred to Raikoke as a "mixed event" and now refer to it as a single-source, volcanic, event. We removed mention of the hypothesis that the "secondary plume" may have contained smoke. Finally, we updated the manuscript to contain a thorough examination of Raikoke's smoke classifications to demonstrate that the Raikoke aerosol was better described as sulfuric acid and to support our conclusion that this event fell outside the scope of this method. While the manuscript has undergone substantial revisions, we provide responses to Dr. Fromm's comments below.

K21 have made substantial and necessary changes to their manuscript.
We appreciate Dr. Fromm's recognition of the effort that went into these changes.

The root of my concerns is that their method of inferring particle composition is based wholly on the spectral slope of aerosol extinction in the visible to near-IR and that they provided no independent observational proof that wavelength-dependent aerosol extinction in this range unambiguously carries particle-composition information. K21 appropriately present the two main elements that might distinguish smoke from sulfates in their imprint on vis-NIR extinction spectra, absorption and particle-size distribution (PSD). However, a priori establishment of these two qualities is not satisfactorily made. The role of absorption is presented in the theoretical realm, embodied in Figures 3 and 5. But the confounding effect of PSD is given inadequate attention. If there is to be a single, algorithmic relation between aerosol extinction and spectral slope, it must be established that volcanic sulfates and wildfire smoke have characteristic, systematic differences in PSD. If there is no such systematic difference between smoke and sulfate, then absorption vs. scattering will rule and there should then be an unambiguous systematic difference on the spectral slope between clearly homogeneous smoke and sulfate populations. The evidence provided in K21's tabular results shows that this is not the case. If in fact PSDs between smoke and sulfate are substantively different, then this factor should be presented as an empirical determinant on the spectral-slope construct (Figure 3). However, to my reading, K21 do not make that case.

We believe that we are in agreement with Dr. Fromm on some of these points. To aid in the clarity of this response we will break the response into individual points.

1. "...their method of inferring particle composition is based wholly on the spectral slope of aerosol extinction in the visible to near-IR and that they provided no independent observational proof that wavelength-dependent aerosol extinction in this range unambiguously

carries particle-composition information." If Dr. Fromm is suggesting that UV-Vis/NIR spectra collected in the SAGE wavelength range are unable to identify smoke "fingerprints" in a manner comparable to an FTIR instrument (such as ACE-FTS) then we are in agreement. However, aerosol scientists have a long history of making scientific inferences using otherwise incomplete information. This has been done for inferring composition based on lidar observations as Ansmann and Ohneiser recently have, separating aerosol and clouds in the SAGE data, and separating aerosol and PSCs in SAM II, SAGE II, SAGE III/M3M, SAGE III/ISS, POAM, and CALIPSO. CALIPSO makes similar inferences for determining aerosol types (e.g., smoke, sulfate, dust, etc.). This results in the researcher applying his or her expertise and making educated inferences on the system. Inferring composition or physical properties is nothing new in aerosol science and that's what was presented in this paper. However, we have updated the text to highlight the limitations of this methodology and enhanced the discussion on potential misclassifications of the Raikoke event's spectra.

As demonstrated within this paper, smoke particles have a significantly different refractive index from sulfuric acid, which significantly alters the spectral slopes. How this change in spectral slope is interpreted was explained in the manuscript and is the point of Fig. 3.

If Dr. Fromm is suggesting that UV/Vis/NIR spectra are incapable of classifying small particles as either smoke or sulfuric acid then we disagree and the case-study analysis is supportive of our conclusion that this is possible. We updated the manuscript to reflect this perspective and help the reader understand that this method is presented as applicable within the context of smaller volcanic eruptions. Raikoke is no presented as an example of where/why/how this method is not applicable to eruptions that produce large particles.

2. "If there is to be a single, algorithmic relation between aerosol extinction and spectral slope, it must be established that volcanic sulfates and wildfire smoke have characteristic, systematic differences in PSD..." While we did not make this claim, and our analysis is not dependent on this assumption, we did discuss the difference in mode radii of background sulfuric acid aerosol and smoke particles (section 4 of the manuscript). We discussed how small amounts of BC substantially change the spectral slope (section 5, and Fig. 5). Further, we discussed how these changes in spectral slope (due to smoke) could be misclassified as large sulfuric acid particles. We provided an example (lines 213-216) of how *small* smoke particles (10% BC) could be misclassified as *large* sulfuric acid aerosol particles (≈200 nm). The manuscript underwent major edits to provide this perspective. As stated above, interpretation of these results require informed scientific inferences much like interpretation of other aerosol observations.

3. "...absorption vs. scattering will rule and there should then be an unambiguous systematic difference on the spectral slope between clearly homogeneous smoke and sulfate populations. The evidence provided in K21's tabular results shows that this is not the case." We apologize, but we have to disagree with this point. The data presented in Table 3 as well as Figures 6, 7, 10, 11 clearly demonstrate a distinct difference in the behavior of the spectral slope for the different case study events: this is unambiguous and incontrovertible. The different behavior was accentuated as the extinction coefficient increased. While we agree that there were misclassifications, such misclassifications were not unexpected and we do not see this as justification for discarding the entire method (no need to throw out the baby with the bathwater). To our point, as shown in Table 3, the algorithm correctly classified sulfuric acid and smoke plumes in the case study events >99.5% and >81% of the time, respectively.

Dr. Fromm clarified his point in offline conversations. His concern was that this table, by itself, does not communicate the fact that much of the Raikoke "smoke" events were misclassifications. While the reader should understand from the text that misclassifications existed, we updated this table to indicate where misclassification occurred, updated the caption, and pulled the Raikoke event data into a separate table. Further, we enhanced the discussion of misclassifications after the Raikoke event by expanding section 7.4. Finally, the manuscript underwent a major revision to present this method as applicable to pyroCb events and small/moderate volcanic eruptions...not larger eruptions such as Raikoke (we present Raikoke as an example of a volcanic event that does *not* fit within the confines of this method and include a lengthy discussion to demonstrate that Raikoke aerosol is better classified as sulfuric acid and *not* smoke). We hope this makes the results of this study more clear for the reader.

A more specific major concern I expressed in the original review relates specifically to the PSD issue. I pointed out that Thomason et al. (2021) (T21) showed that even within the superset of presumed volcanic sulfates, vis-NIR extinction spectra vary widely among diverse volcanic events. Although K21, in their response to my review, rightly argue that T21 did not consider possible composition mixes, the large range of spectral extinction variation among plumes that are almost certainly exclusively volcanic (i.e. the plumes from tropical eruptions) speaks to a very wide constraint on pure-sulfate PSD. Raikoke falls within the range of these tropical sulfate plumes (T21's Figure 8, below), hence it is a considerable challenge to argue that there is a motive for questioning the Raikoke-year's composition mix considering T21.

The major finding of T21 was that volcanic eruptions of various sizes induced changes in extinction coefficient and extinction ratio that were extremely well behaved. Defending that manuscript is outside the scope of the current manuscript and the review process. However, T21 and K21 are not in disagreement with each other, rather they corroborate each other. The point of confusion may be that the current study is not as simple as suggested; we did not look solely at the spectral slope (or extinction ratio) to determine composition. As explained within the manuscript we performed a statistical analysis that involved pre- and post-eruption/pyroCb extinction coefficients and spectral slope. It was the overall change in both of these parameters, relative to background conditions, that determined the classification type. This is distinctly different from the T21 methodology. While T21 and K21 are similar in many respects, one cannot simply take the findings of T21 to build a framework for interpreting K21.

We hope we do not misunderstand Dr. Fromm's comment, so we'll provide another response in hopes of "hitting the target." We agree that the size of the sulfuric acid aerosol formed by the Raikoke eruption fall outside the scope of this method. Our classification method works very well on pyroCb smoke and small/moderate volcanic events, but Raikoke clearly falls outside the paradigm of this method. The paper was updated to reflect this. Raikoke is interesting in and of itself and we must admit that the various observations of the scientific community are not easily reconciled (we discuss this toward the end of the manuscript).

K21 assert in their response to my review that T21's analysis has a different motive and construct as compared to K21, but extinction ratio and a spectral slope derived therefrom are inextricably linked to the same underlying determinant: small particles yield both a large extinction

ratio and spectral slope.

As explained above, the two studies, while similar in many respects, are fundamentally different and the results of the current study cannot be constrained to a framework built upon T21. The case-study events in the current manuscript are supportive of the proposed methodology. However, we also recognized within this manuscript that large eruptions, such as the Raikoke eruption, pose a distinct and substantial challenge because of the amount of SO_2 they release and the resulting particle sizes. We made a major shift in how this method is presented and communicated in this manuscript: the classification method works well for pyroCbs and small/moderate volcanic eruptions, but Raikoke is presented as an event that falls outside the paradigm of this method. We added discussion to the manuscript regarding possible explanations for the observed behaviour (including misclassifications for Raikoke data). None of the outlined scenarios perfectly account for the observations and some fall outside the current scientific dogma. Therefore, as discussed in our conclusions, there may be something unique about this event that is not fully understood.

We believe Dr. Fromm's point here (after offline conversations) may be that Raikoke was indistinguishable from the wildfire events and that there may be other events that behaved in a similar manner (e.g., Pinatubo, and the 1990 Kelut eruption). Dr. Fromm suggested that this is supported by the results of T21. We agree that Raikoke 2019 produced large sulfuric acid particles (we provide size estimates as well as go into lengthy discussion on why smoke classifications do not fit) that resulted in smoke classifications. We note that Pinatubo and 1990 Kelut may exhibit the same behavior (we cannot state definitively because the method was not developed for SAGE II). We hope the expanded discussion better resolves this issue and makes the situation better understood with the reader.

It is not logically clear why K21 make an exception for Pinatubo, when inferred volcanic sulfate PSD covers a quasi-continuous range from small to large (T21). Indeed Raikoke is situated deeply into the large-particle side of the PSD spectrum according to T21. Moreover, Thomason (GRL, 1992) characterized Pinatubo aerosols as the norm for volcanic clouds such that the "new mode" found within SAGE II extinction spectra was labelled "new" and hypothesized to be a transitional mode in volcanic sulfate-particle growth. To the extent that pyroCb smoke explained the Thomason (1992) "new mode," its relevance to the current manuscript is central to my assessment that vis-NIR extinction spectra are inadequately constrained for particle-composition inference. To wit, had K21's spectral slope algorithm been applied to Pinatubo, it is likely that SAGE Pinatubo measurements would have been classified as smoke and pyroCb smoke measurements would have been closer to sulfates.

Dr. Fromm's claim that our algorithm would classify the Pinatubo plume as smoke is in agreement with the original and revised manuscripts. From lines 256–259 of the reviewed manuscript: "At this point we must reiterate the caveat that this holds true only for small or moderate eruptions and would not be applicable to larger eruptions such as the 1991 eruption of Mt. Pinatubo. Eruptions that inject large amounts of ash and SO2 into the stratosphere, like Pinatubo, have been observed to rapidly produce extinction ratios indistinguishable from water clouds and presumably smoke." From lines 610–614 (emphasis added) of the reviewed manuscript: "during large-scale volcanic eruptions (e.g., Pinatubo with VEI 6), the *probability for misclassifying the large sulfuric acid particles, or ash, as smoke is high*. Under this scenario, massive amounts of ash and SO2 is injected into the stratosphere, resulting in the formation of large sulfuric acid aerosol particles (e.g., > 500 nm). This would result in a spectrally flat extinction spectrum, which *effectively mimics the behavior of smoke in the slope analysis. Usage of this method in either of these conditions*

is questionable." Since we have already provided explanation within the original and revised texts we do not provide further change here. However, the most-recent revision reframed the applicable bounds of this method and we think this will help the reader understand that this method is not to be used for large events like Raikoke and Pinatubo.

The point I made in my first review about T21 was simply that their illustration in Figure 8 was sufficient to caste large doubt on any attempt to derive a binary choice between stratospheric smoke and sulfates. T21's "perturbation" 520/1020 nm extinction ratios are all over the place, from less than background to the upper limit: unity. Roughly half the plumes have increased extinction ratios (more negative slopes, in K21 parlance) w.r.t. background, the remainder have smaller extinction ratios (less negative slope). Raikoke is in a class with three other plumes with decreased extinction ratio, each created by a tropical volcano and no question as to contributory smoke. The different motives and methods of T21 vis a vis K21 are not germane. The patterns within T21 Figure 8 map directly to K21. The fact that T21 only consider the background and peak perturbation in Figure 8 serves to clarify my point, not diminish it. The T21 perturbation state represents extinction signals at absolute and relative maximum, a state of minimum uncertainty in the signal. Hence it is a demonstration of the optimal circumstances for assessing both microphysical evolution and composition. Hence, T21 Figure 8 shows that there is no firmer footing for questioning Raikoke's composition mix than for questioning that of any of the other three plumes in Raikoke's class.

The manuscript underwent major revisions (mentioned above) and we believe all of Dr. Fromm's points within this comment have been addressed. We apologize for not addressing Dr. Fromm's comment explicitly, but the aforementioned changes should satisfy this comment.

In summary, the construct presented in T21 bears two lessons for K21. One is that the Raikoke-era plume is indistinguishable from several presumably sulfate-only plumes. The other is that it is necessary for an algorithm aimed at discerning composition (or blends) needs to go beyond a smoke/sulfate binary outcome.

These are interesting points. However, we must raise the cautionary note that the frameworks of the two studies are fundamentally different and one cannot determine composition based solely on extinction ratios. To Dr. Fromm's point, there are lessons that can be learned from both studies and expanding beyond a "binary" outcome would make for an interesting future study.

At the risk of "beating a dead horse," I offer one other argument related to T21. I took the example of Kelut (February, 1990) to illustrate how this volcanic cloud could suggest a smoke blend, according to K21's construct. A basic rendering of 1-micron extinction and 525/1020 extinction-ratio as a function of time are shown below. The data are version 7 SAGE II. One representative retrieval altitude (19 km) makes up this example. The extinction data are limited to profiles between 0-20°S latitude.

Next is an analysis akin to K21's Figure 4: color ratio as a function of 1020 nm extinction. It is clear that this distribution of background color ratio to perturbation color ratio is more like the two smoke examples in K21 Figure 4 and the two sulfate examples. Undoubtedly, Kelut's inarguably sulfate plume at 19 km would have been in the smoke category within the K21 construct (modified as it would be for SAGE II wavelengths). Not shown, but evident at other stratospheric altitudes

up to 21 km, is this same pattern. At higher altitudes it morphs to a more K21 sulfate pattern. This is reminiscent of the altitude variation of smoke/sulfate proportion systematics presented in K21. To me this is further evidence of the weakness of the SAGE data and the K21 construct for composition inference.

We appreciate Dr. Fromm's comment and do not mind a dead horse being beat. On the contrary, we appreciate Dr. Fromm taking the initiative to perform this test. However, while the aforementioned changes to the manuscript should satisfy Dr. Fromm's concerns, we must state that the analysis Dr. Fromm presents is not representative of the method within the manuscript. The major problems with Dr. Fromm's analysis are:

1. Applicability of this method to SAGE II has not been tested. While we expect a similar methodology may be used on SAGE II data, this was not the focus of this manuscript; therefore, Dr. Fromm's comparison is outside the scope of this manuscript. Here is an example of a potential complicating factor: depending on altitude and latitude SAGE II had 2–3 applicable aerosol channels while SAGE III/ISS has 6–7 applicable channels. Further SAGE III/ISS has a 1.5 μm channel (SAGE II does not have this channel), which has proven highly valuable in this analysis. Therefore, direct application to SAGE II has not been tested and is not necessarily straight forward, making Dr. Fromm's example case tenuous.

2. The method Dr. Fromm used in creating these plots was not the same as what was presented in this manuscript. Dr. Fromm used a simple extinction ratio and did not consider how these ratios changed relative to background conditions (we also used spectral slope, not extinction ratio, though this is a minor difference). On the other hand, our method invoked altitude and latitude-dependent statistics of background conditions on an event-by-event basis and carried out a statistical analysis to determine the criteria for background, sulfuric acid, and smoke classifications. If Dr. Fromm's point is that a single extinction ratio (sans additional channels and a statistical characterization of background conditions) cannot be used to infer composition then we agree completely. However, that is not the method used in this study making Dr. Fromm's comparison invalid.

In response to referee suggestions, K21 invoked additional satellite data to qualify their findings. For instance, they showed TropOMI SO2 in concert with a SAGE profile (Figure 1) deemed to sample sulfates. This was very helpful. However, it would help the reader if K21 would refer to Figure 1 and report on the algorithmic result for this profile (as they did of another example shown in Figure 19). At a glance, the 520/1550 extinction ratio at peak extinction is quite small (6.7) in comparison to the two sulfate distributions in K21 Figure 4 (10 being the minimum) and much closer to the two smoke examples.

We thank Dr. Fromm for the insightful suggestion. The aerosol in this profile was classified as predominantly smoke. While this is interesting, we chose not to include this information within this figure for 2 reasons: 1. the figure appears early in the manuscript before any of the methodology was presented, so adding smoke classifications may add confusion for the reader, 2. we provide a lengthy discussion on why Raikoke falls outside the paradigm of this method and why there were so many smoke classifications; we do not see how adding classifications to this figure adds value, 3. a new figure was added (Fig. 20 in the most recent revision) that demonstrates that smoke classifications were present *very* soon after the Raikoke eruption. In short, the manuscript has

undergone a major revision and we do not see the proposed change as being terribly beneficial to the reader within the context of this new version.

They also introduced ACE-FTS gas and aerosol information. In this endeavor, they cited Chris Boone of the ACE-FTS team for a still novel aerosol-infrared spectrum construct for aerosoltype determination. While this led to improvements in the manuscript, in my opinion K21 drew a conclusion regarding the secondary and primary Raikoke plumes that fall short of convincing.

We find this comment to be confusing two reasons. First, the ACE-FTS smoke identification method is a robust technique that relies on IR spectral signatures of smoke and this technique has been used in multiple publications. Further, criticism of the ACE-FTS identification methodology is outside the scope of this manuscript and this review cycle and it is not clear to us how criticism of the ACE-FTS technique applies to the current manuscript. Second, we stated explicitly that "While Fig. 18 supports our finding that smoke was in the stratosphere during this time period and within the latitude range of our study, *it does not provide confirmation for all of our smoke classifications*" (lines 539–540, emphasis added), and "...ACE-FTS data showed no spectral evidence of smoke above ≈20 km" (line 634), and "...we conclude that it is reasonable that the aerosol layers we classified as smoke during the Raikoke time period *were composed of larger sulfuric acid aerosol* and we were unable to find definitive support for smoke at altitudes above 20 km." (lines 642–644, emphasis added). In this context, we find Dr. Fromm's comment to be confusing. However, we updated one of these conclusion statements to help clarify our conclusion: "...we were unable to find definitive support for smoke at altitudes above 20 km" now reads as "we were unable to find *any* definitive support for smoke at altitudes above 20 km." (emphasis added to highlight the change.)

While we failed to understand the motivation for Dr. Fromm's comment we hope that the major changes we made to the manuscript will satisfactorily resolve any ambiguity for the reader: Raikoke resulted in large sulfuric acid particles and serves as an example of how/when/why Raikoke falls outside the applicable bounds of this method.

For instance, K21 state that they find no ACE evidence of smoke above 20 km, using Figure 18 as an illustration. Absent though is any discussion of whether this conclusion is also informed by other ACE profiles not shown. Did K21 find, for instance, no CO or HCN enhancements above 20 km? Did they find a substantial number of ACE profiles with such enhancements up to 20 km? If so, that would bolster their conclusion.

This comment is not representative of the conclusions of the manuscript. Finding more evidence for stratospheric smoke would go against the conclusions of the manuscript. As stated in response to a previous comment, we conclude that while smoke was present in the stratosphere during this time period, the Raikoke aerosol layers were composed of large sulfuric acid particles (lines 634, 642–644 of the reviewed manuscript). Since the conclusion is in agreement with Dr. Fromm's position we do not recognize any changes to be made solely for this comment. However, we did make major modifications to the manuscript that hopefully resolve any ambiguity.

Specific to Figure 18, there is a concern with respect to the profile in which the ACE aerosol-infrared spectrum approach showed a smoke signal at 20 km (Figure 18 j,k,l). The smoke marker is at 20 km, above any CO or HCN enhancement, but centered on a huge SO2 enhancement. At

a glance this appears to be confusing at best, erroneous at worst. Given that this profile is at 24N, 34E (situated on the western side of the Asian Summer Monsoon anticyclone, it may be a case of upper tropospheric biomass burning signal below the SO2 peak, and a volcanic plume above. It should be examined more closely before a final determination, but as it stands, it is not a compelling foundation for K21's conclusion that smoke was observed up to 20 km.

Defense and extensive explanation of the ACE-FTS method for identifying smoke (per Boone et al., 2020) is outside the scope of the current manuscript and this review cycle. That said, there seems to be confusion regarding this method and we apologize for this confusion (and we thank Dr. Fromm for pointing this out). These profiles were identified as candidate smoke profiles because CO and HCN markers showed enhancement. Therefore, Dr. Boone used the aforementioned smoke identification method to determine the highest altitude where smoke was detected (i.e., using spectral residuals to locate smoke signatures within the residuals). In his analysis, there were traces of smoke at 20 km. We have updated the manuscript to include this nuance.

Regarding the stratospheric Raikoke-season plume that ascended to 22+ km at low latitudes, what K21 refer to as the "secondary" plume, it is evident that the smoke angle and this anomalous plume height are a prime motivation for this paper; it is highlighted in the abstract. The original and revised K21 manuscript provide an unclear motivation to pursue this as smoke assisted. In the Introduction, K21 state: "The working hypothesis was that this secondary plume consisted, at least in part, of wildfire smoke." The "was hypothesized" part implies that the idea preceded K21 yet no citations are given. Also in the Introduction, they cite 3 papers on the subject of diabatic smoke-plume lofting, but none of these involve the Raikoke-season plume. I did not find any attribution of this hypothesis to the Raikoke secondary plume" cited papers. If I missed it, the authors are asked to provide the earlier sources. Otherwise, the "working hypothesis" should be claimed as their own.

We thank Dr. Fromm for this comment. Dr. Fromm is correct that the stated hypothesis was our own working hypothesis. The papers cited in reference to diabatic lofting were included as a reference for how smoke can loft air masses; no pretense was made to these papers being from the Raikoke time period. That said, the manuscript underwent major changes and reference to this secondary plume self lofting because of smoke have been removed. We believe that the tone and message of this manuscript has fundamentally changed and resolved this comment.

Related to the above point, the original manuscript was criticized for implicating "2 major wildfires" (Abstract) as contributors to stratospheric aerosols in 2019. The criticism was that the manuscript relied on citations that provided only vague and inconsequential support for the wildfire/pyroconvection angle. I argued that the fires and pyroconvection in 2019 were not "major" in comparison to a normal pyroCb season (while acknowledging that stratospheric smoke injections were an annual occurrence). Yet the current manuscript retains these arguments. My challenge to K21, to establish a strong argument for wildfire/pyroconvection influence, went largely unmet. That challenge remains and is an essential component for convincing the reader that fires such as those in 2019 could independently bolster the argument for a wholesale pollution of the stratosphere in the Raikoke season. I have a suggestion, to look at a pyroCb season such as 2018 with SAGE data. In that year an equal number of pyroCbs occurred in the northern hemisphere as 2019, but in the absence of a volcanic cloud and the demonstrably huge smoke plumes in the NH in 2017 or the Australian plume in 2019/20. If a sizable smoke presence were to be found, it would

bolster their argument that a normal pyroCb season (such as 2019's) could present a blend worth considering in the Raikoke year.

Defense of peer-reviewed publications cited within our manuscript is outside the scope of the current manuscript and outside the scope of this review cycle. We find Dr. Fromm's comments to be confusing as we did *not* claim that the 2019 wildfires were responsible for "a wholesale pollution of the stratosphere in the Raikoke season." We did hypothesize that smoke contributed to the lofting observed in the SAGE and CALIOP data, but support for this hypothesis was not found (as noted in the manuscript). This comment fails to recognize the conclusions of this study that "While Fig. 18 supports our finding that smoke was in the stratosphere during this time period and within the latitude range of our study, *it does not provide confirmation for all of our smoke classifications*" (lines 539–540, emphasis added), and "...ACE-FTS data showed no spectral evidence of smoke above ≈20 km" (line 634), and "...we conclude that it is reasonable that the aerosol layers we classified as smoke during the Raikoke time period were composed of larger sulfuric acid aerosol and we were unable to find definitive support for smoke at altitudes above 20 km." (lines 642–644). Our only interpretation of Dr. Fromm's comment, within the context of the reviewed manuscript, is that we are in general agreement with Dr. Fromm on this point and the wording within the abstract may be confusing.

While we continue to mention the 2019 wildfires in the most recent version (the one Dr. Fromm is now reviewing), we no longer include the hypothesis of smoke contributing to self lofting and the message of the manuscript has undergone major revisions. We believe, within the context of this revision, that Dr. Fromm's comment here has been satisfied.

Regarding Dr. Fromm's challenge: we see no need to meet this challenge as it would appear that the conclusions of this study (that stratospheric smoke, while present, was not a significant source of aerosol during this time period and the Raikoke plumes were composed of large sulfuric acid aerosol) is in agreement with Dr. Fromm's suggestion.

Overall Interpretation of the Reviewer's Comments

First, we are grateful to Dr. Fromm for his time and consideration in reviewing this manuscript. We view Dr. Fromm's comments as helpful in improving this manuscript. The overarching response from Dr. Fromm seems to indicate that sections of the manuscript were confusing and the messaging was "off". We apologize for this confusion and we made major revisions to the manuscript to address these issues.

Dr. Fromm suggested that many of the cited works were problematic (not that we incorrectly referenced these works, but the published results that we cited were problematic). We note that all of these works passed peer review, some of which Dr. Fromm reviewed, and the assigned editors chose to proceed with publication. Therefore, we view citation of these works as appropriate. We cannot address criticisms of these previous studies as that falls outside the scope of this peer-review cycle. Therefore, we retained reference to these published studies.

The manuscript underwent major revisions, part of which is intended to better convey the message that this method is intended for use with small/moderate volcanic events (and pyroCbs) and is not applicable to larger events (Raikoke serving as the demonstration of why/how it fails). We updated Table 3 (now broken into 2 tables with asterisks indicating misclassifications) inclusion of an additional figure (Fig. 20), as well as an expansion of section 7.4 to include more discussion on why large sulfuric acid particles are more likely than smoke as well as a discussion of possible scenarios that may have led to these misclassifications.

2 Response to Reviewer #2

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

Dr. Fromm's recommendations to consider the problem in a multi-facet way and give more ground to their analysis by simultaneously using complementary datasets was followed by adding a section based on the exploration of two single SAGE III/ISS events and co-located ACE-FTS events

It is true that Dr. Fromm suggested we present corroborative data from other instruments to support the assertion that smoke was present in the stratosphere during the Raikoke time period. The reviewer is correct that this was achieved by presenting data collected by the ACE-FTS instrument. It is unfortunate that only 2 SAGE/ACE-FTS events were coincident with each other; that is why the comparison was limited to 2 events. However, as demonstrated in the manuscript, smoke was identified in additional ACE-FTS profiles that covered NH latitudes (from tropics to Arctic). This additional data provided the requested support for our assertion that smoke was present within the stratosphere. Perhaps this was not clear in the manuscript so the text has been updated to include this.

"...based on the exploration of two single SAGE III/ISS events and co-located ACE-FTS events, an analysis lacking convincing conclusions (except that this requires more investigations) and bringing more uncertainties than strong demonstration of their method."

We disagree. The ACE-FTS analysis provided an unambiguous demonstration of the presence of smoke within the stratosphere. This does *not* indicate that the stratosphere was dominated by smoke and we do not make that claim in the manuscript. The comparison specified by the reviewer was limited to 2 coincidences because these were the only 2 coincident observations as stated above. We agree with the reviewer that more coincident profiles would be beneficial, but this is not available.

This new section looks fastly written: the language becomes approximate (e.g.,"the slope" becomes sometimes a concept per se), the discussion concerns for a significant part the existence of a lot of uncertainties, and the amount of "personal communications" replacing a solid analysis supported by published studies is quite high.

We apologize, but we do not understand the reviewer's meaning behind "'the slope' becomes sometimes a concept per se". The spectral slope is a mathematical product that, based on a statistical analysis of background conditions, is used for classification.

We must disagree with the reviewer's assertion that use of personal communication is taboo. The personal communications we referenced were with research groups that have published methodologies for determining aerosol composition (Chris Boone of the ACE-FTS group) and particle size distribution parameters (Felix Wrana). We asked these groups to use their methodologies to estimate mode radius and composition for example profiles. The papers that describe these methods were reference in the manuscript as well. It appears that the paper already has what the reviewer suggests (i.e., solid analysis supported by published studies).

Unfortunately, the authors did not take the time to integrate Dr. Fromm's recommendations to the whole study, missing the opportunity to reap the benefits of such a revised approach: work made in sections 2-7.3 remains roughly unchanged (except some addition like Fig. 5 expliciting the "theoretical considerations" and providing a somewhat more generalized treatment of this ideal case)

Our revised manuscript did incorporate recommendations from Dr. Fromm if they were applicable to the current study. To our knowledge there is only 1 instrument capable of making an unambiguous identification of stratosphere smoke (ACE-FTS) and data from that instrument was incorporated in the first revision. TropOMI was included to demonstrate the presence of SO_2 over North America (as requested). We also included the CALIOP vertical feature mask as part of the original analysis. Any additional satellite observations would provide a tenuous (at best) classification of the composition; therefore, addition of these data sets would not be terribly elucidative. If the reviewer has specific suggestions for inclusion then we would be appreciative for the discussion.

"...admissions multiply about the weakness of some elements of the discussion ("We recognize that" is used four times)."

The reviewer is correct that we were transparent about the limitations of this study and the assumptions that went into the theoretical understanding. This helps the reader understand the applicability of this method (strengths as well as limitations).

More satellite datasets could have been used (Reviewer #4 make a comment in this sense) but this possibility was apparently not investigated carefully. We agree that it is important to understand the strengths and weaknesses of each individual instrument, as mentioned by the authors in L. 98-99 and in the reply to Reviewer #4, but using other datasets to confront results obtained by the instrument under investigation looks the best way to draw strong conclusions about these strengths and weaknesses.

This statement is not accurate as the reviewer failed to notice that 2 additional satellite datasets were added to the analysis for the revised manuscript. As noted above, this was considered and we agree with the reviewer that additional datasets strengthen the analysis/conclusions and better define the study's strengths and weaknesses. Indeed, this is why ACE-FTS and TropOMI data were added to this analysis. However, we chose a limited dataset based on data availability and the applicability to this study.

The fact that the refractive index of carbonaceous particles shows a high variability, but that the authors built their study on an arbitrary choice of this refractive index remains a weak point of the discussion.

This is not correct. The preliminary theoretical work used a series of compositions (refractive indices) to demonstrate general behavior patterns (i.e., to demonstrate from a theoretical perspective that the ensuing analysis is sound). We then evaluated the spectral slopes of four case-study events to determine whether the behavior (as predicted from theory) was observed in reality. This was done without assumption of particle composition; hence there was no "arbitrary choice of this refractive index."

The fact that the same concept of "spectral slope" is used in two different ways (around Fig. 3 and in the implementation of the real case (§5)) is also quite disturbing for the reader.

We apologize to disagree, but the spectral slope was used consistently throughout the study. Section 5 of the manuscript discusses the spectral slope's sensitivity to changes in composition. Perhaps this was not clear in the text; the text was updated to elucidate this nuance for the reader.

Concerning Table 3, the authors didn't take into consideration the Reviewers' concerns, e.g. related to the high values of some smoke [sulfuric acid] fractions in cases of pure volcanic [wild fire] events although they affect, if we consider the total number of valid spectra, quite a lot of measurements. Yet, this would be a discussion or a clarification worth to clearly identify the strengths and weaknesses of their method.

We agree with the reviewer that this was an opportunity to discuss the strengths and weaknesses of this method. Such a discussion was included in the original and revised manuscripts. Section 5 of the manuscript was included in the revised manuscript to aid the reader in interpreting this. As <0.5% of the enhanced volcanic spectra were misclassifications (34 spectra out of >6400) we assume the reviewer is referring specifically to the wildfire events (≈19% of the enhanced wildfire spectra were deemed misclassifications). These misclassifications were discussed from lines 403–425 of the revised manuscript, so it is unclear what additional discussion could be included. Given that the submitted manuscript already contains the requested discussion, we did not make any changes regarding this point.

The criticism expressed by several referees with respect to the "theoretical considerations" around Fig. 3 led the authors to provide to several warnings (L. 218, l. 230-238), that this theory is just interesting to get some insight into the problem but is actually not used in the following and basically unusable. This is basically confusing for the reader and could have been brought forward in a more appropriate way to show the relevance of their approach.

We appreciate the reviewer's comment. We did not claim that this theoretical framework is not usable, but the reviewer's suggestion and point is well made. This figure and the corresponding sections provide a theoretical foundation upon which the reader can build an understanding of the physical processes within this study. Contrary to being "unusable" this is highly useful for interpreting the results and *was* used to interpret the results throughout the manuscript. However, the subsequent analysis was not constrained to this theoretical paradigm (i.e., the observed extinction spectra were used without forcing them into a specific framework). We updated the text to clarify the intent of this figure.

As a result, the structure of the paper looks like a compilation of poorly coordinated blocks, first with a theoretical model based on many [not usable] assumptions and actually "not used" in the following, followed by an analysis including several difficulties that could have been better understood by using insights from other datasets, followed by a fast analysis of some "SAGE" events (actually: "SAGE III/ISS", and not "SAGE", which is another instrument) where the authors show additionnal difficulties and challenges leading to the conclusion that it can be interesting for future studies. All these together form a patchy ensemble missing the opportunity of a strong paper.

Therefore, I recommend that the authors fully revise the paper and mainly take the time to

really implement in an integrated and consistent way the recommendations formulated by the referees through the whole study. Either this can be made in a succesfull way and the authors will be able to reorganize the paper and make each element a logical piece in an integrated whole; or the combination of insights from different instrument show that the original argumentation is poorly grounded, and in this case, the authors should limit their conclusions to what is valid, or in the worst case, unfortunately, recognize that their original analysis was not well-founded.

Text within the manuscript was updated to provide better continuity between sections. The manuscript underwent major changes to better communicate this work.

Overall Interpretation of the Reviewer's Comments

First, we are grateful to this reviewer for his time and consideration in reviewing this manuscript. We view these comments as helpful in improving this manuscript. The overarching response from this reviewer seems to indicate that continuity between sections of the manuscript was not clear. We apologize for this have updated the manuscript for clarity sake.

The reviewer provided several comments about the limited scope of our analysis. The manuscript was updated to indicate this is because of the limited applicability of additional datasets and the extremely limited number of coincident observations available.

The reviewer provided several comments regarding missed opportunities for discussion within the manuscript. Many of these "missed opportunities" were already in the manuscript, and it was not clear as to how these discussions were insufficient. Therefore, we did not address these comments in a specific manner.