We are very grateful to the reviewers, Daniele Visioni and Anonymous Referee #2, as well as for Pasquale Sellitto for providing their insightful and constructive comments on the manuscript. We appreciate input and think that the paper benefited from the suggested changes.

Below, the answers to the reviewers' questions and comments are provided. In order to make the text more distinguishable, we highlighted the reviewers' and Dr. Sellitto's comments in **bold** and authors' answers in blue font.

Reviewer 1: Daniele Visioni

This paper combines measurements and model simulations to assess the outcomes of the 2018 Ambae eruption. It is an excellent study, and more of this kind are needed to understand the shortcomings of climate models in simulating volcanic plumes evolution and the radiative effects of the increase stratospheric sulfate. I think this paper is great for ACP, it is scientifically robust and novel. I have multiple suggestions to improve the text in places (especially the readability in English) and I have listed them below, but I don't think this is generally an obstacle for the publication of a relevant advance in our scientific knowledge, so after these suggestions are considered, the manuscript can be published promptly.
Abstract
Line 2: "These effects are more noticeable after..."

Corrected.

Line 4: no comma after studies

Corrected.

Line 4: "Besides" is not the right word here. I'd actually rephrase the entire concept using: "There have been several studies, where a volcanic eruption plume and the associated radiative forcing were analyzed using *either* climate models *or* data from satellite measurements: however, studies combining both models and measurements are rare."

During the manuscript revision, it was decided that this phrase in the abstract was inaccurate. We revised it to: "There have been several studies, where a volcanic eruption plume and the associated radiative forcing were analyzed using climate models and/or data from satellite measurements. However, few have compared vertically and temporally resolved volcanic plumes using both measured and modelled data."

Line 11: remove "the", leaving "for most latitude bins"

Corrected.

Line 12: remove ":"

Corrected.

Line 17: "to the" instead of "through"

Corrected.

Introduction

Line 24: "is" now well established.

Corrected.

Line 41: this is a nice overview. However, it does not mention quiescent degassing and non-explosive eruptions that are also an important source of sulfate for the UTLS, see for instance Pitari et al. (2016)

Thank you, we agree that those sources are important in the formation of the stratospheric aerosol layer. It is worth mentioning though, that talking about "moderate" (VEI 2-3) volcanic eruptions we also include non-explosive (VEI<4) eruptions. However, we agree that it is important to add quiescent degassing as a source. Therefore, we changed the sentence to "Although these sources along with quiescent volcanic degassing are undoubtedly important, the large scale changes to the stratospheric aerosol layer are primarily driven by moderate and large volcanic eruptions which emit sulfur dioxide (SO_2) directly into the upper troposphere lower stratosphere (UTLS) region (e.g. Kremser et al., 2016; Pitari et al., 2016, and references therein)."

Line 44-45: just to be precise, but it's not the models themselves, it's the underlying simulations without proper sulfate emissions. For instance, in Schmidt et al. (2018), they compare simulations with and without volcanic emissions in the same model.

We agree, it was not precise wording. We changed the sentence to "According to Solomon et al. (2011); Haywood et al. (2014); Schmidt et al. (2018, and references therein), it has been shown that climate models' simulations that neglect forcing from volcanic eruptions since the year 2000 tend to project a faster rate of global warming for the first 15 years of the 21st century than the simulations including this volcanic forcing."

Line 57: comma after volcanoes.

Corrected.

Line 58: "in" instead of "of"

Corrected.

Line 62: "...the last decade, although it did..." (just fixed the punctuation)

Corrected.

Section 2

Line 127: remove comma after "highlight"

Removed.

Section 3:

I'm a bit confused by the structure of the sections: wouldn't it be better to have a Section 2 with a description of both the measurements (OMPS, SAGE III) and possibly the ECHAM simulations, and then move to the results? it would be a bit clearer to the reader.

The structure of the sections has been changed in accordance with your remark and recommendations from referee #2.

Line 153: "Ångström" is missing an s(but it has all the correct accent marks, which usually nobody bothers with!)

Corrected.

Line 159: no comma after "Both"

Corrected.

Line 164: no "the" before the dates

Removed.

Line 168: this is my ignorance, being a modeler, but how can there be negative Extinction values? Is it an error in the retrieval algorithm (and if that's the case, why not remove them?). Or something else? If it can be briefly explained, it should be.

We agree that intuitively the negative aerosol extinction coefficient seems to be erroneous. However, it is not exactly an error, rather a feature of the retrieval algorithm and are statistically possible. Damadeo et al. (2013) discuss this problem for SAGE II instrument. In their paper, the authors mention, that negative extinctions can occur in the situations when the extinction value is low and signal-to-noise ratio is small. They also state that removing them would introduce a positive bias to the extinctions. We haven't found a literature reference for SAGE III, however, Robert Damadeo (pers. com. 2019) recommended us not to filter negative values for SAGE III exactly for the same reason, which was stated in the lines 168-169 of the original manuscript. We revised them by adding " We did not filter negative Ext_{869} because this would bias the comparison (see Damadeo et al. (2013) for details)."

Line 178: "observed" instead of "seen"

Corrected.

Line 187: I see what the authors mean, but "put" is not the right word. If the original observations where spatially non uniform, and the authors obtained a uniform grid out of them, then it's technically a 2D interpolation.

We see, our expression was confusing. What we actually mean is the binning rather than interpolating. The paper text has been corrected.

Section 4

Line 202: remove "the" before height

Removed.

Line 206: a periodicity, or periodic signal

Changed to periodic signal.

Section 5

Line 216: remove "the" before altitudes

Removed.

Line 217: "However" instead of Although

Changed.

Line 229: northward and southward

Corrected.

Section 6:

Again, it's a bit weird to have Fig. 3 with the ECHAM results before a description of the simulations. I'd have most of this section much earlier in the text.

We changed the structure of the paper to account for your comments and those of the reviewer #2

Line 247: for two reasons.

Corrected.

Line 261: the phrase "Furthermore, the radiative cloud fractions less than 0.2 and a solar zenith angle less than 70° were required" doesn't make sense to me.

This is a very relevant objection because the radiative cloud fraction is not needed to filter the OMI data when the retrieval algorithm assumes an SO_2 profile with a center of mass in the lower stratosphere. Therefore we recalculated the SO_2 burden without a constraint for this variable. We still excluded all SO_2 data that were measured at solar zenith angles above 70° as it is recommended by the OMI manual: "Volcanic SO_2 data from all rows of the OMI, with the exception of rows affected by the row anomaly, can be used. As with the PBL SO_2 data, it is best to use retrievals from scenes with SZA<70° " [citation from the manual:"OMSO2 README File v1.3.0 Released Feb 26, 2008 Updated: June 16, 2016" `https://so2.gsfc.nasa.gov/Documentation/OMSO2Readme_V130_20160616.pdf`] We rephrased the sentences to: "Only the measurements obtained at solar zenith angles less than 70° were used."

Section 6.2: this section is explained in an excellent way, but I have a doubt that I can't seem to find in the text: what's the horizontal extent of the SO_2 in these measurements? I.e. when the authors say "Finally, the SO_2 mass of the entire grid per batch are summed up to obtain the total SO_2 burden.", what's the horizontal span of the grid that contains, say, 99% (or 95%) of the SO_2 used in the estimates you present for the volcanic eruption? This is potentially an interesting point to include.

We are very grateful to the reviewer for this comment. However, it was very challenging to give an estimate on the position of the longitudes that contain 95% of the SO_2 mass inside the plume, and sadly we are not sure if there is a definite answer.

We provide an estimate using the following technique, first, we averaged the SO_2 mass (the threshold was applied beforehand) over all batches, see the left panel of Figure 1 of this document. To get the horizontal extent of the plume, we subsequently averaged over all latitudes (right panel in the Figure 1). Finally, we moved the the boundaries of the outer longitudes towards the center until the sum of the SO_2 mass between the boundaries reached 95% of the total SO_2 mass. Using this approach, we may say that 95% of the SO_2 mass was detected between 152.9°E and 142.8°W.

Line 326: "vertical" instead of "altitudinal"

Corrected.

Line 335: most aerosols are

Corrected.

Line 339: "blob" is not a very technical word... Maybe just "area"? (also in line 376)
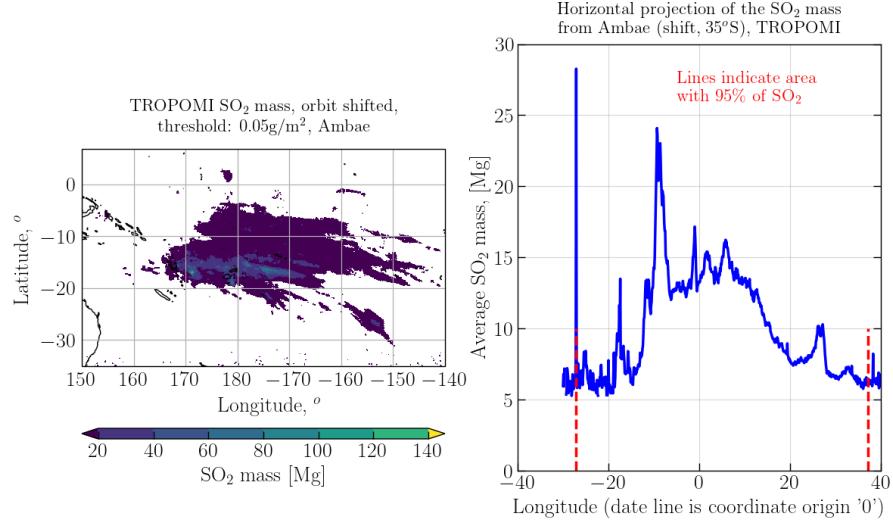
Figure 1: Left: average of the SO_2 mass over all batches. Right: the time averaged SO_2 mass was averaged over each latitude to estimate the expansion of the plume. The red dotted lines mark an area that contains 95% of the total SO_2 mass.

The text has been changed to "A disagreement is seen, however, around the 19.5 km altitude in November, where OMPS-LP data show an increased extinction not present in ECHAM simulations." to account for a comment of reviewer #2

Section 7

Line 360-361: just a comment, but I'm also quite amazed at how well the plume is reproduced in ECHAM! This is a great result.

Thank you! We are also very satisfied with it.

Line 368: "with" observations

Corrected.

Line 374: observed instead of "seen"

Corrected.

Line 382: is the SO_2 amount big enough to produce a noticeable w* increase produced by the heating? Can the authors give some information on the stratospheric temperature perturbation? (the nudging ensures that this doesn't translate in additional w*, of course, but can we actually see an effect on stratospheric temperatures with 0.4 Tg-SO_2?)

In the ECHAM model only atmospheric waves with large wave numbers are nudged. The residual vertical velocity should change. However, as the reviewer speculates, the eruption was too small to show a clear signal on the residual vertical velocity. We compared our simulation to a control simulation without
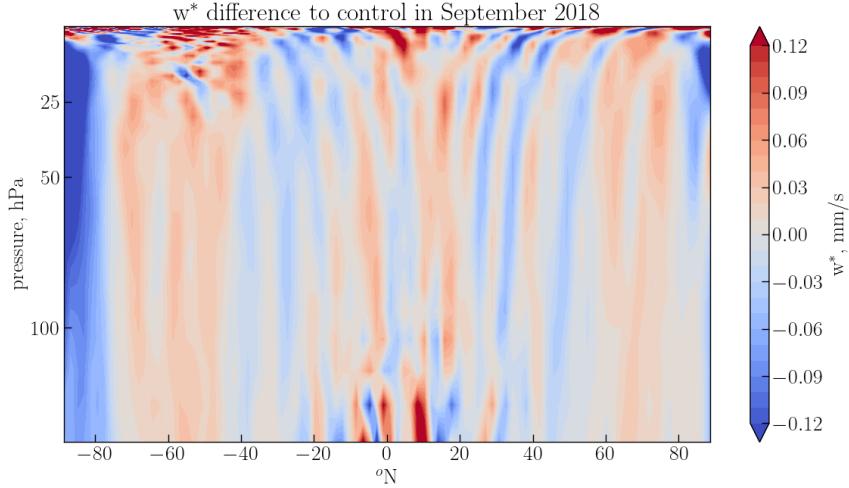
Figure 2: Difference in vertical velocity (w*) in September 2018. The difference was obtained by subtracting the control run (without the eruption) from the run with the eruption.

a clear result.

The Figure 2 shows the anomaly of the residual vertical velocity as a difference to the control run, a nudged simulation without volcanic eruption.

Line 384: "The" Ambae eruption

Corrected.

Figure 6: there might be some problem with the legend: the (869) part is missing.

Thanks for drawing our attention to it. This might have happened during the manuscript uploading, now the plots are in png-format, so the problem is resolved.

Line 422: "the" reader's attention

Corrected.

Line 433-434: how do you estimate the 0.13 W/m2 value? Doesn't look like it's the value from the green line only. Is it an average between all curves? Some? An explanation is warranted.

This is a value from the blue solid line, calculated using the "original" Hansen formula. We change the sentence to "For the particular Ambae eruption studied in this paper, using Hansen's formula, we estimate the tropical radiative forcing caused by an increase in stratospheric aerosols to be about -0.13 W/m^2."

Conclusions

Line 446: "observed" instead of "seen"

Corrected.

Line 447: "the" tropics

Corrected.

Line 449: "global aerosol model" is not really precise? I'd suggest using "GCM with a coupled aerosol microphysical model"

We change the text to "from a GCM with coupled aerosol microphysics".

Line 453-456: This phrase is a bit confusing. I've tried to reword it below a bit "For the first eruption ECHAM underestimated the strength of the plume as well as the time by which it reaches 20.5 km of altitude, whereas for the second eruption the modeled plume reaches higher altitudes about two to three weeks earlier, and the plume lives longer while being slightly weaker overall at those altitudes"

Following the comments from anonymous reviewer #2, it was decided to revise the discussion without the phrase about first eruption, as it is too small to be traced uniquely.

Line 459: "the" aerosol radiative forcing

Corrected.

Line 461: the time "evolution" is a better term

Corrected.

Line 466: "the model" with no "s". Can't assume the same for other models...

Changed to "the model gives".

References

Pitari, G., Visioni, D., Mancini, E., Cionni, I., Di Genova, G., & Gandolfi, I. (2016). Sulfate aerosols from non-explosive volcanoes: Chemical-radiative effects in the troposphere and lower stratosphere. Atmosphere. https://doi.org/10.3390/atmos7070085

Damadeo, R. P., Zawodny, J. M., Thomason, L. W., and Iyer, N.: SAGE version 7.0 algorithm: application to SAGE II, Atmos. Meas. Tech., 6, 3539–3561, https://doi.org/10.5194/amt-6-3539-2013, 2013.

Reviewer 2: Anonymous Referree #2

In the manuscript "Changes in stratospheric aerosol extinction coefficient after the 2018 Ambae eruption as seen by OMPS-LP and ECHAM5-HAM" by Malinina et al. the Ambae eruptions in 2018 are investigated using multiple ($\approx$ 6) satellite measurement datasets and a model simulation. The first 10 pages (Sections 2-5) focus on the OMPS-LP extinction retrieval, data quality, and result in an OMPS-LP extinction climatology and a section focusing on Ambae. The second part (section 6) on the model simulation introduces four more satellite data sets, derives SO_2 mass injection time series for Ambae, presents the model setup and the results. At this point I started wondering what this study was about and what would be the key result(s). The discussion section compared the OMPS-LP and ECHAM model results and additionally introduced an estimate of the radiative

forcing. Comparisons to previous studies on Ambae and references to studies observing similar effects when comparing measurements and satellite observations are missing. In my opinion this study contains a lot of material that merits publication, but the material need to be sorted, the results should be put into reference of existing knowledge from previous studies and the key message(s) and conclusion(s)should be worked out and stated explicitly. Hence, I'd recommend for publication after a major revision. Please find major and minor comments below.

Major comments:

Scientific objective: To me the scientific objective of the study is not clear. It seemed that the Ambae eruption was studied as an end in itself and certain aspects, such as assessing the quality of the OMPS-LP extinction retrieval algorithm, estimation of mass injection time series, and the radiative forcing estimation were just some by-catch.

The main objective of the paper is to investigate the influence of the Ambae eruption on the state of the tropical stratosphere and how different the perturbation of the stratosphere is seen by observations and the model. When presenting comparisons of the observed and modelled data it is necessary to know the quality of the observational data. As no assessment of the quality of OMPS-LP retrievals at IUP Bremen was published before, this needed to be done in the framework of the current study. In the course of the comparisons it was found out that the model results are very sensitive to the initialisation conditions, i.e. the amount and altitude of SO_2 injected into the stratosphere. To make this initialization correct, a careful investigation of the injected SO_2 mass and injection altitude was necessary. To ensure the traceability of the study the description of the methods needed to be included into the paper. Finally, the radiative forcing is an important and easily understandable and comparable measure of the importance of a particular volcanic eruption for climate and it should not be omitted in such a study. Thus, in our opinion all presented elements of the study are necessary to make the study complete and the results reproducible.

Please make clear why did you perform the ECHAM simulations? Which new aspects did you learn from the ECHAM simulation that the measurements did not provide? Did you learn anything from the differences between observation and model? Do you have recommendations for improved volcanic plume simulations?

There is a large scientific community performing model runs and investigating the development of volcanic plumes in the stratosphere as well as the implications of SO_2 injections into the stratosphere on the climate. Studies comparing the observational and modelling data for isolated small-scale eruptions are, however, quite rare. Model intercomparison studies (e.g. Clyne et al., 2021) revealed strong differences between the results of the evolution of the volcanic cloud of different models. Aerosol microphysical processes are highly non-linear and e.g. differences in transport can result in quite different particle distribution and size. Similarly, differences in microphysical processes between

the models can have a strong impact on simulated forcing. Therefore, comparing model results with satellite products can lead to improvements of the model results, and in turn, model results can also help to improve satellite products.

In addition to a general scientific interest in observational and modelled data comparisons for an overall assessment of the model capabilities, the usage of the ECHAM model in our study enables us to calculate the radiative forcing resulting from the Ambae eruption and to assess the validity of the simple approximation suggested by Hansen et al. to estimate the radiative forcing in a limited latitude region (i.e. not globally as originally suggested) based on the stratospheric aerosol optical depth.

There were indeed some lessons learned in course of the model simulations and comparisons. Namely, a high importance of accurate estimation of the injected SO_2 mass and injection altitude as well as a need for a nudging of a free running climate model with external information on the atmospheric dynamics (ECMWF ERA 5 in our case). We observed the Ambae aerosol plume at 20.5 km appeared several weeks earlier in ECHAM data although ECHAM was previously considered to have too weak vertical lifting in the not nudged version. The reasons for this behavior are currently under investigation.

We revised our introduction to explain better the motivation for the model data intercomparison study and the expected benefits. We also revised our discussion and conclusions by adding our model runs findings.

Paper structure: In section 6 four new satellite data sets (MLS, OMI, OMPS-NM and TROPOMI) are introduced to derive the mass SO_2 injection time series and injection altitude. These data set description are scattered over all sub- and sub-subsections and distract from following the line of arguments that should lead to the model setup. I'd recommend to introduce all instruments and data sets at the beginning in an own section on instruments and data sets. I assume section 6.2 should be section 6.1.2.Please also consider a methods section. E.g. in section 6.1.1 and section 6.2 the method to grid OMI/OMPS-NP and TROPOMI SO_2 data seem identical. The subsection on the radiative forcing in the discussion section 7 belongs into the main part of the paper.

We agree that the paper needs re-structuring. Based on the reviewer's suggestion and on the suggestion from Daniele Visioni, we changed the manuscripts outline to

1. Introduction
2. Instruments and model
2.1 OMPS-LP
2.2 OMPS-NM
2.3 MLS
2.4 OMI
2.5 SAGE III/ISS
2.6 TROPOMI
2.7 ECHAM
3. Observational data:

It is important to note here, that the MLS SO_2 profiles as well as OMI, OMPS-NM and TROPOMI SO_2 concentrations aren't a part of a new satellite dataset we acquired ourselves. We just used the existing products to make an assessment of a plume height and SO_2 burden. We've adjusted the text to make it more clear.

Discussion: A presentation of the key finding(s) and a discussion with respect to existing knowledge from previous studies on e.g. Ambae, volcanic eruptions into the UTLS in the tropics, and simulations of volcanic plumes is missing. Please see suggestions in detailed comments.

We thank the reviewer for proving valuable suggestions on the improvement of the connection of our study to the existing knowledge. We did our best to implement the reviewer's suggestions. For details please see our answers to the comments below.

Detailed comments:

l31-33: Please provide a reference.

A reference to Kremser et al. (2016) was added.

l35: The reference is from 2011 and does not cover the bush fires in 2019 mentioned in the text.

Indeed, the paper from 2011 covers only fires in 2009. By the time of submission of our initial manuscript to ACP, to our knowledge there were no peer-reviewed papers available about Australian Bushfires 2019. In the mean time, this changed, thus, we added the reference to Khaykin et al. (2020).

l37: Please add a reference for pyrocumulonimbus, e.g. Fromm et al., 2010

Added.

l46-52: Here, multiple simulation studies investigating volcanic plumes are listed. However, the scientific questions that are addressed and the findings are not mentioned. Investigating volcanic plumes is not an end in itself.

Indeed, each of the cited papers focuses at specific scientific questions ranging from validation of the models, through investigations of perturbations in the atmospheric state, to implications for stratospheric ozone and climate. However, as our paper is not meant to be a review paper we are not able to discuss the scientific focus and findings of each particular paper we cite. This would blow up the introduction, overload it with a lot of details not related to the focus of our study and make the whole paper much more difficult to read. The interested reader can check the cited publications directly. We acknowledge the fact that

the mentioned papers are just a subset of available publications not pretending to be comprehensive. For that reason, we added "e.g." before each citation block.

l81: Please add a separate section for all the instruments and data sets used through-out the paper: OMPS-LP, OMPS-NP, SAGE III, MLS, OMI, TROPOMI.

The paper is restructured as discussed above.

l100: "... from all altitudes from 290 to 1000 nm ..." Please provide the altitude range here.

As we mention in the following sentences (lines 104-105 of the original manuscript), the OMPS-LP altitude range varies with latitude and season; however, the altitude range from 5 to 80 km is constantly covered. We think that providing this information at the place suggested by the reviewer would be redundant.

l125/126: What was the highest retrieved extinction? Please add this information. Why did you chose this thresholds? Please justify. Did you take further measures to filter out ice clouds? What about volcanic ash? Does volcanic ash affect the sulfate aerosol extinction retrieval?

The reviewer asks valid, interesting and relevant questions. A lot of those questions scientists from the limb-scattering community asked themselves and each other for a couple decades, however, fully satisfactory answers haven't been found so far. Below, we briefly summarize the state of the art on those topics.

The highest retrieved extinction is 4.0978×10^{13} km^{-1}. However, we don't believe that this information would be of any interest for users. This value occurs on the 28th of June 2017 at 08:40UTC at 26.8°N, 66.7°E at 10.5 km, which is most likely to be in a thick convective cloud.

The threshold to reject clouds is selected empirically to keep as much as possible of the aerosol extinction and reject as many clouds as possible. The trade-off is determined by the potential application of the data set. For applications, where it is more important to get rid of as many clouds as possible and single high aerosol peaks are not that important, a rather conservative value of 0.002 km-1 is used. This value is based on the results from Bourassa et al. (2010), where the Ext_{750} after the Kasatochi eruption were not exceeding 0.0015 km-1. We used this threshold when we previously showed our OMPS-LP data (e.g., Malinina, 2019). For the investigation of an isolated volcanic eruption, as, for example, the Raikoke eruption 2019 (Mauser et al, 2020) or in this study, a higher threshold is necessary as we are interested in preserving all increased aerosol values. As we investigated the plume propagation at rather high altitudes we do not rate a possible contamination by clouds as a crucial issue. Thus, for the Raikoke case the threshold was set to avoid loosing any increased extinction value above the tropopause, which resulted in the value of 0.1 km-1. We also used this threshold for Ambae.

We also want to apologize, as while preparing the answer to this comment, we've noticed that we used the plot with the old cloud cut-off threshold (0.002 km^{-1}) for Figure 2 of the original manuscript. We've changed the Figure in

this version of the manuscript to the correct one. It resulted in a slight change of the section.

We haven't implemented any additional measures to filter ice clouds. As the reviewer might be aware, this has been a serious and extensively discussed issue in the limb-scattering community, but generally no satisfactory solution has been found so far. Different groups use different approaches, e.g. NASA OMPS-LP (Chen et al., 2016) uses the altitude derivative of the color-ratio approach. For the latest OSIRIS v7 product, the altitude derivative of the color-ratio as well as extinction information were used to detect clouds. However, all those approaches aren't perfect. They all tend to remove high extinctions. We believe, that the cloud filtering using a threshold is as good as any of them. Though there are still some minor artifacts and/or thin clouds present in the product, at least the important changes in extinction connected to some events won't be filtered out.

A similar issue exists with respect to volcanic and/or forest fire ash/soot. Basically, any change in stratospheric aerosol composition leads to a change in refractive index, which in turn changes the retrieved extinction. For example, Bourassa et al. (2019) have done some studies on this topic. They have found that small changes (from 1.427 to 1.41) in the real component of the refractive index result in changes in the extinction coefficient of about 30%, which is a "rule of thumb" error for most limb-scatter aerosol products. Large scale changes result of course in larger errors. Unfortunately, there is no established way so far to determine the aerosol composition from currently available space borne limb-scatter measurements. Thus, for now we are forced to stick with a fixed refractive index.

It is also an important note that, generally, ash particles are relatively large and sediment quickly out of the stratosphere usually already during the first days after an eruption, although some very fine ash particles can remain in the stratosphere for longer (Vernier et al., 2016). Thus, in general our studies assessment of the plume should not be affected by ash.

Concluding, the reviewer's questions are highly relevant, however, answers to most of them are work in progress for the whole community, and each of them requires at least an own dedicated paper to be addressed.

l131-145: Please shift to instrument and data set description section.

Done.

l139: ..is a near infrared photodiode...

Corrected.

l145: At which wavelengths are the SAGE III extinctions provided?

NASA provides Ext at 384.2, 448.5, 520.5, 601.6, 676, 756, 869.2, 1021.2 and 1543.9 nm. We added this information to the manuscript text.

l155-157: Why would the OMPS instrumental uncertainties not impact a comparison with SAGE III? I'd rather think that instrument uncertainties would contribute to differences between both instruments. Please provide an explanation.

We agree that this sentence was misleading, we've changed the text to "Firstly, SAGE III is an independent data set; thus, possible OMPS instrumental issues (e.g. scattering angle dependency) will be revealed by the comparison, which would not be the case when using other OMPS products".

l163: Which Ångström exponent did you use? What is the larges difference between 830-900 nm?

This must be a misunderstanding. The Ångström exponent does not need to be used here. To avoid confusions, the sentence has been reworded as: "Another advantage of the comparison with SAGE III is the same measurement wavelength. Both, OMPS-LP and SAGE III provide measurements at 869 nm, so no conversion of the aerosol extinction to any other wavelength needs to be done."

l166: Since you used SAGE III solar occultation and OMPS-LP solar scattering measurements, what was the minimum time difference between the profiles that were compared?

With the applied geographical criteria, the minimal time difference was 1 hour 47 minutes 37 seconds and the maximal difference was 22 hours 7 minutes 38 seconds. We added this information to the text.

l177-180: It would be more interesting to know the reasons for the differences and not only who studied it. Please add a short summary of the reasons.

It looks like our descriptions was confusing. We provide the key findings from Rieger et al. (2018) in the next sentences. We adjusted the paragraph and changed the word "Thus" to "According to this study" to make it more clear.

l180-181: Is this a finding by Rieger et al. (2018)?

Yes, to make it clearer in the text, we changed the word "Thus" to "According to this study".

Figure 1 and 170-174: Can you comment on why OMPS compared to SAGE-III systematically overestimates the extinction in extratropics? Did you filter out polar stratospheric clouds (PSCs) at high latitudes in winter time? Did both data sets include PSCs?

There was no screening for the clouds other than filtering by 0.1 km^{-1}, so yes, PSC might be present in both datasets. However, it's not the reason for the observed systematic differences. The main reasons are summarized in the lines 180-182 of the original manuscript: "Thus, the most important sources of errors in limb retrievals arise from the uncertainly assumed aerosol loading at the reference tangent altitude as well as the unknown aerosol particle size distribution parameters. The latter factor mostly affects the high latitudes where the viewing geometries are close to forward and backward scattering."

l186-187: If you have to average the extinction data rather depends on what you want to study. I.e. if you are interested in maximum plume height or detailed transport and conversion processes you'd rather not average the data.To create Fig. 2 actually no pre-gridding as described here is necessary. It probably even may introduce artefacts. Which vertical bin size did you use?

We absolutely agree with the reviewer that the necessity of data averaging is determined by the study objectives. In the next sentence, we explain that we provide two types of level 3 products. One of them (monthly averaged) aims to provide a general climatology, another one (10-days averaged) is more suitable for plume studies. Because of a relatively sparse sampling of OMPS-LP, a plume tracing with single measurements is not feasible. The OMPS-LP repeat period is about 8 days, so in order to see global distribution of aerosols, one needs to make some sort of average. However, since the measurements do not occur at the same spot each day, one needs to provide some sort of regular grid.

Of course, one can draw a time series plot simply using independent measurements without any binning. However, such a plot would look much noisier and possibly hiding regular patterns. It is a rather common practice in the scientific community to show monthly zonal mean time series when analyzing multi-year time series of atmospheric species. We do not see why this widely used approach would be unsuitable for our manuscript.

There is no vertical averaging needed to create OMPS-LP level 3 product as Ext is retrieved on a constant 1-km grid (see section 2 of the original manuscript). We added the word "regular" there to make it more clear. We also changed the ending passage highlighted by the reviewer to: " products were binned onto a regular geographical grid with 2.5° latitude and 5° longitude steps. Since the retrieved product is provided on the regular 1-km grid, no vertical averaging is needed."

Figure 2: It is not clear if here an average of your level3-product is shown, or if these are real monthly zonal 30° averages.

We show level 3 product averages everywhere except Fig. 1 of the original manuscript. We changed the caption to "Monthly mean aerosol extinction coefficient (Ext_{869}) distribution as a function of time and altitude. The values were obtained by binning and zonally averaging OMPS-LP monthly level 3 Ext_{869}."

There are reoccurring vertical white stripe patters in winter time. What is causing these stripes?

The stripes result from the seasonal variations of the aerosol extinction coefficients with maximum values in the winter months and much more pronounced structure in the northern hemisphere.

Why is the data cut at 16 km? In the extra-tropics this is well above the tropopause and missing the lower stratosphere.

We extended the altitude range in the Fig. 2 to 13.5 km for the extra tropics. We chose 13.5 km because it is an average altitude of the tropopause at 30° latitude.

We also want to apologize, as while preparing the answer to this comment, we've noticed that by mistake we used the plot with the old cloud cut-off threshold (0.002 km^{-1}). We've changed the Figure in the manuscript to the correct one. It resulted in minor change in the text of the section though.

201: Please add reference to Vernier et al., 2011 for the aerosol tape recorder effect.

The reference was added.

l205-11: I find this explanation confusing and to some extent misleading. Do you mean the annually reoccurring white stripes in Fig. 2 here? This pattern is visible at all altitudes at the same time in winter. This does not look like QBO to me.

The reviewer is correct, the white stripes occur not because of QBO at all, this is the annual oscillation which was described in the lines 208-211 of the original manuscript. However, we see that our explanations were confusing. We reformulated the paragraph by adding the years of the QBO occurrence. We also added the phrase about " yearly re-occurring lighter coloured stripes" into the sentence about annual seasonality.

Please mark or zoom into the mentioned QBO pattern, because I cannot see it in Fig. 2.

QBO manifests itself as a modulation of white stripes at high altitudes with higher values in 2013, 2015, 2017 then in the intermediate years. It is best seen in panel (b) of the figure with dark blue bands in the first five years. We added the years into the manuscript.

What causes the yearly changes in stratospheric aerosol loading? Or is this annual pattern rather an instrument artefact? Please explain.

The main driver of the seasonal oscillations is the Brewer-Dobson circulation transporting aerosols from the tropics. It has its maximum strength in the winter months in the northern hemisphere. There might be, however, some contribution from the changing scattering angle causing instrumental artifacts.

l219: "...averaged over longitudes.." Do you mean zonal means here? Or did you restrict these averages to a certain longitude range?

Yes, here we mean zonal means. The text was changed accordingly.

l222-224: How large was the increase in extinction after the April eruption? Is this increase significant? In Fig. 3 I cannot see any increase after the April eruption. At 18.5 km a slight increase to the north of Ambae is already visible before the eruption. Between 30-40S the extinction remains constant until June. Please make clear which increase you mean and provide numbers/factors for/of the extinction increase that match with what is shown in Fig. 3.

We agree with the reviewer that the increase in the first week it is rather a speculation. We changed the text and now start discussion with the increase in May. "Already in the first week after the eruption a small increase in Ext869 is seen around the Ambae location, this increase cannot be, however, uniquely attributed to the Ambae eruption and can be caused by the transport of the aerosol from preceding events."

l232: Does the plume really vanish by mid-October? Fig. 2 shows that the plume is still there, but at higher altitudes. Please rephrase.

We think the word "vanish" was used without context. The sentence where it was used is "By mid October 2018, the plume starts to vanish around the equator and continues to weaken with time". So yes, reviewer is correct, the plume doesn't vanish. We changed the sentence to : "By mid October 2018, at

18.5 km, the plume shows a clear reduction around the equator and continues to weaken with time. "

242,251: Why did you use MLS data to obtain vertical information of the plume? MLS has a significantly coarser vertical resolution than e.g. CALIOP and OMPS-LP. Why didn't you use CALIOP and OMPS-LP to derive information on the injection altitude?

It is true that CALIOP and OMPS-LP have a better vertical resolution in the UTLS. However, neither of those instruments provides SO_2 profiles. The only approach we see to roughly estimate which altitude a volcanic cloud reached is through linking it to the increase in extinction/backscatter ratio due to ash. However, ash cloud is not the same as SO_2 cloud. Additionally, using that technique, one can estimate the highest reached altitude, but not a vertical profile. As we mention in the text, the derivation of the SO_2 injection altitude from nadir instruments is related to high uncertainties. Thus, to our knowledge vertical profiles of SO_2 from MLS, even with their disadvantages, are the best option.

l248-250, Fig. 4b: Showing the OMI/OMPS-NP SO_2 mass injection time series for the July eruption too would provide valuable information. Comparing both approaches would allow for an estimate of the uncertainty of this approach for the April eruption due to the use of different instruments. Are there also spatial gaps in OMI/OMPS-NP data after the April eruption?

We repeated the SO_2 mass calculation for the July eruption using the OMI/OMPS-NP data set and obtained a maximum SO_2 mass of 0.54 Tg (panel b in Figure 3). This was surprising because it gives a higher maximum SO_2 mass than the TROPOMI estimate although the OMI/OMPS data exhibit large data gaps. Please note that the self-defined grid for the first eruption contains latitudes up to 45°S whereas the grid for the second eruption only reaches to 35°S in order to be comparable with the TROPOMI analysis. Finally, there are, unfortunately, still data gaps in the OMI/OMPS data set after April.

l253: How deep was the SO_2 injected into the stratosphere? How does the injection altitude profile look like?

While preparing the original manuscript, we hesitated adding the plot with MLS profiles, since we do not derive this product ourselves, but only use the existing data. However, we decided to add the MLS SO_2 profiles as well as their description into the Appendix A1 to make it more clear.

Section 6: I'd recommend to separate instrument descriptions from the method description, the results, and their discussion. In particular the descriptions of the grid and method (l262-272 and 284-291) seem redundant.

This comment has been addressed by a general restructuring of the paper as suggested by the revier.

l258: Please describe briefly what the OMI row anomaly is.

The OMI row anomaly is a phenomenon that affects the quality of the radiance data for all wavelengths in a specific viewing direction of the instrument. It is believed to stem from a damage in the isolation that blocks part of the
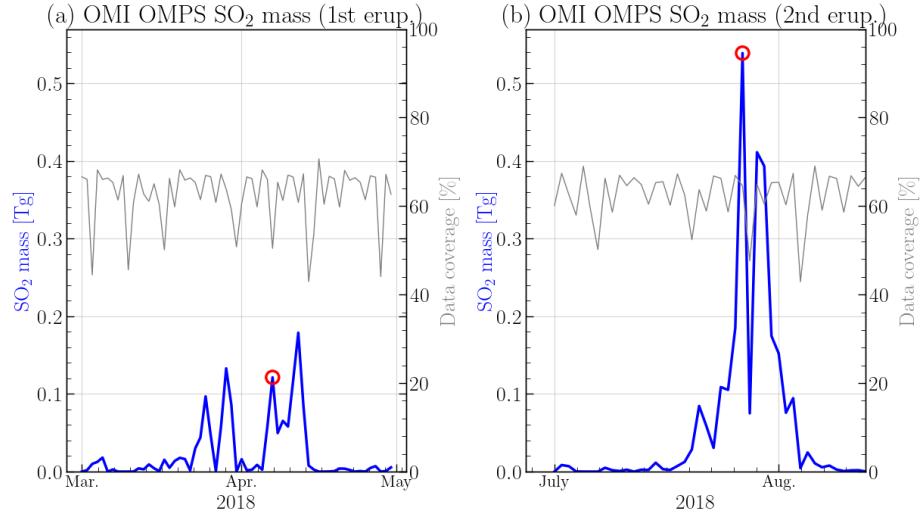
Figure 3: SO$_2$ mass for both April and July eruption, calculated from combined OMI/OMPS-NM data.

instrument's field of view (Levelt et al., 2018). We added the explanation to the text.

l262: When introducing a threshold to distinguish background from volcanic signal, please provide information on the OMI sensitivity towards SO$_2$. Please state why you selected 0.05 g/m2 as a threshold.

The threshold should discriminate the volcanic SO$_2$ signature from the natural SO$_2$ background and was found empirically. Furthermore, the threshold was introduced to filter out negative SO$_2$ data that were referred to the manual as negative retrieval noises: "When estimating the SO$_2$ loading from a volcanic plume within a given domain, it is recommended that only OMI scenes or pixels exceeding a certain threshold (e.g., 1 DU) be included in the calculation. This helps to filter out occasional negative retrieval noise." [OMSO2 README File v1.3.0 Released Feb 26, 2008 Updated: June 16, 2016]. The threshold of 0.05 g/m2 that we used corresponds to 1.75 DU. According to Krotkov et al. (2016), the OMI detection limit for SO$_2$ is ca. 0.2 DU.

l263: Why do you convert g/m^2 to g/m^2? Please clarify.

Sorry, the question is unclear. The SO$_2$ vertical column densities are provided in Dobson units and have to be converted to g/m^2. However, we understand that our formulations cause the confusion, so we changed the order of the sentences to make the explanations more clear.

l258,266,277-278,286: I think a little bit of explanation of the assumptions made for the centre of the SO$_2$ mass altitude would be helpful. This information is distributed over the text it take some

text forensics efforts to understand that these differences introduce some uncertainty. How sensitive is the result on the assumed mass altitude?How much would the result change if 7 km were assumed for TROPOMI? Please add to Fig. 4.

We've added the results for the 7 km SO_2-profile to the plot in Figure 4. To exclude artifacts from TROPOMI that appeared at mid southern latitudes, we reduced the size of the grid for the analysis of the July eruption. The new grid ends at 35°S instead of 45°S. Furthermore, we've changed the definition of the 24h-batch in order to keep the grid as long as possible in the center of the area that is covered by the satellite in one day.

We added the Figure 4 and it's short discussion to the Appendix of the manuscript.
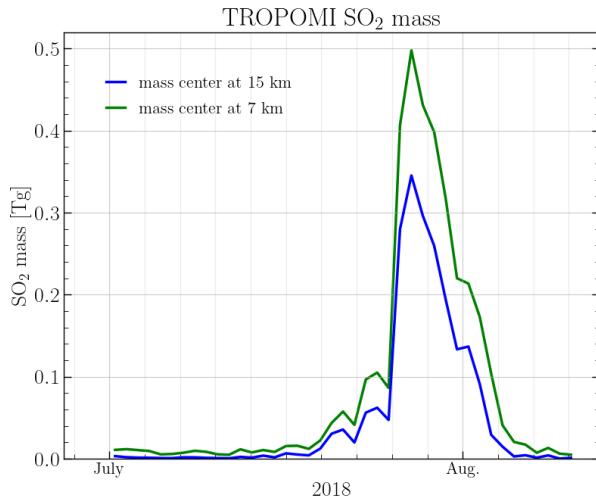
Figure 4: SO_2 mass for the 7-km and the 15-km-SO_2 profile.

l275: Please provide information on much of the self-defined grid was covered by TROPOMI before applying the threshold to provide a reference value.

TROPOMI covered the self-defined grid nearly completely. Before the threshold was applied, 96-97% of the self-defined grid contained data that could be used for analysing the SO_2 product which assumed a TM5 model profile. An even higher coverage of 97-98% was obtained when SO_2 products were used that assumed box profiles (15 km and 7 km), because a different data filtering method was applied. We added the information about TROPOMI data coverage to the text.

l287: Why did you only consider column densities less than 1000 mol/m2?

Vertical column densities larger than 1000 mol/m2 were excluded because

they were considered unrealistic and erroneous. We added this information to the text.

l308: Please explain, what is TM5 model?

The TM5 is a global chemical transport model that provides a daily forecast of SO_2 profiles (Theys et al., 2017). We added the explanation to the text.

l311: What does MECHAM stand for? What is the difference to ECHAM?

The reviewer is right, we could have mentioned this more clearly. MA stands for the middle atmosphere version. MAECHAM5 is an extension to a pressure level of 0.01 hPa compared to 10 hPa of the standard version of ECHAM. We changed the text to: "The volcanic eruptions were modeled by MAECHAM5-HAM. ECHAM5 (Giorgetta et al., 2006) is a general circulation model (GCM) which was used in the middle atmosphere (MA) version, a high top model version with maximum altitude at 0.01 hPa (about 80 km)" We also checked the spelling of the model name throughout the text.

l320: ... from Sect 6.1 and 6.2, right? From which data set were the altitudes derived?

There was a mistake in numbering in the original manuscript. The section 6.2 was meant to be 6.1.2; thus, yes, we meant 6.1 and 6.2. The altitudes were obtained from MLS profiles. We corrected the numbering and added the information on the MLS profiles.

l324: Where does the OH field come from, some climatology?

For the purpose of not giving too many details, we gave only a reference to previous papers. MAECHAM5-HAM prescribes oxidant fields of OH, NO_2, and O_3 on a monthly basis, as well as photolysis rates of OCS, H_2SO_4, SO_2, SO_3, and O_3. OCS concentrations are prescribed at the surface and transported within the model. The climatological data were taken from simulations which included a full gas phase chemistry.

We changed the text to: "A simple stratospheric sulfur chemistry is applied above the tropopause (Timmreck, 2001; Hommel et al., 2011). ECHAM prescribes oxidant fields of OH, NO_2, and O_3 on a monthly basis, as well as photolysis rates of OCS, H_2SO_4, SO_2, SO_3, and O_3."

l345ff, Fig. 3: I'd suggest to add a difference plot to show the agreement and regions of largest difference, probably due to the wild fires.

Thank you for this suggestion, however, we have to reject it. The suggested plot would have a low information content showing very large differences in certain space-time points and giving a false impression of a disagreement between the observations and the model. The objective of the our study is to qualitatively assess the strength, distribution and development of the plume as seen by the space borne instruments and the model. At a current technical level of the model no high quantitative agreement in point-by-point comparison can be expected.

l358-359: I doubt that the plume remains at the same geographical location. There is zonal transport. It rather remains at the same latitude band.

We agree that the wording was not optimal. We changed the text to : "While the plume in ECHAM data stays with the time in the same latitude band mostly in the Southern hemisphere, in the OMPS data, it has a C-shape around the equator."

l364: Please describe your internal studies on ECHAM SO$_2$ sensitivity in more detail. What did you investigate? Could these studies provide some kind of uncertainty estimate?

A key issue, which improved the agreement of observations and model results, was to nudge meteorological variables. As ECHAM tends to have too strong meridional transport, this step keeps more aerosols within the tropics. We varied the injection rate, following several estimates from OMI/OMPS-NP and TROPOMI, from 0.36, 0.4 to 0.5 Tg (SO$_2$) for the main eruption in July. We further increased the vertical extension of the injection area from one to three model levels. This increased the injection altitude slightly and changed the aerosol microphysics as the injection rate per grid box decreased. The results of the study show the importance of a realistically simulated transport. However, we don't think that we can use the results to estimate uncertainties.

Figure 5 of this document shows the sulfate burden, the integrated mass, within the tropics (20N to 20S) for April to November 2018. Simulations A to C show results with nudged meteorological values, in simulation D only the tropical winds were nudged until June 2018. Simulation A and B inject into three vertical levels, simulation C and D into one level only.
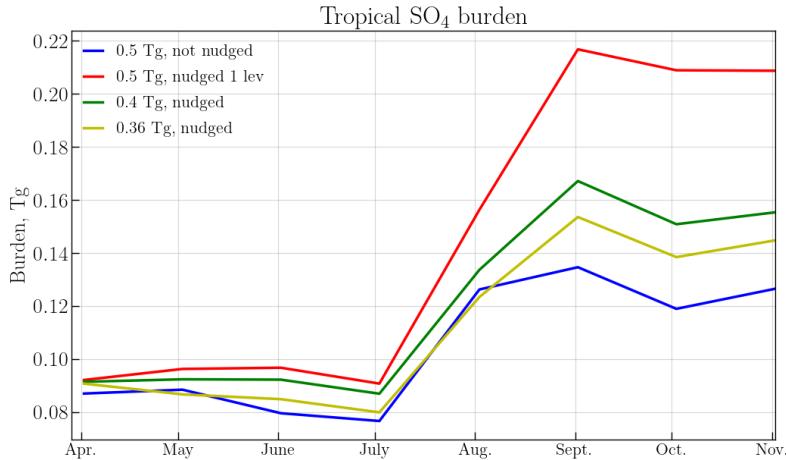
Figure 5: Tropical (20°N to 20°S) SO$_4$ burden in several ECHAM simulations.

We added a figure showing the extinction of ECHAM without nudging to the supplement.

l368: I wouldn't consider ERA5 a pure model product. It has a

substantial amount of measurements assimilated.

We absolutely agree with the statement. In our original text the phrase "a model product as well" referred to the fact ERA5 reanalysis has not only observational, but also a modelling component. However, we agree that this sentence might be misleading. For that reason, we change the text to: "However, the nudging database, the ERA5 reanalysis, has not only observational, but also a strong model component".

Fig. 5: Here I'd also recommend to add a difference panel.

Please, refer to the answer to the comment to l345ff, Fig. 3 .

l374: Actually I think the April plume is nearly invisible. What is the OMPS-LP extinction detection limit? Which changes can be considered significant?

We agree that the April plume is very small and it is stated in the manuscript at several places. Considering tropical observations at 18.5 km altitude, the increase of 10^{-4} km^{-1} as seen in April produces a spectral response which is on the level of the assumed noise, which can be considered as a measure of the detection limit (although purely mathematically a detection for SNR < 1 is possible). The increase of 2×10^{-4} seen after May is well above this limit.

l376: At 19.5 km in November the OMPS-LP maximum is about 1.5×10-3km-1 and 1.0×10-3 km-1 in the ECHAM simulation. Isn't this a significant difference compared to the increase from about 1.0×10-4 km-1 to 2.0×10-4 km-1 at the same altitude for the April plume?

Indeed this difference is significant. Perhaps our phrasing was sub-optimal and resulted in confusion. We changed the sentence to "The plume has the same overall shape and is located at the same time coordinates. A disagreement is seen, however, around the 19.5 km altitude in November, where OMPS-LP data show an increased extinction not present in ECHAM simulations."

Section 7.1: A discussion of the results is completely missing. How well do your results agree with e.g. Kloss et al., 2020; other estimates of Ambae/Aoba SO$_2$ mass injections (`https://so2.gsfc.nasa.gov/omi_2004_now.html`)?

Kloss et al., (2020) considers the Ambae eruption from a completely different perspective putting more focus to large scale effects in both time and geographical extent. As other instruments and a different wavelength of the OMPS-LP data set are used, there are no values, with exception of the radiative forcing, which can be directly compared. The only comparable feature we see is that the maximum of aerosol extinction seen in November 2018 occurs at around 19 km. We include this information in the revised paper text. For other results we see no goal in reproducing findings of Kloss et al., (2020).

Additionally, we do not know any other published peer-reviewed studies which investigated the Ambae eruption from atmospheric point of view. The source suggested by reviewer is a NASA-GSFC website with a side note about the Ambae eruption and doesn't refer to any published sources. The Moussallam et al. (2019) study mentions a paper in preparation on SO$_2$ estimations, however it looks like this paper has not (yet) been published. There was also a twitter

post by a very reputable volcanologist, Professor Simon Carn. However, we find ourselves uncomfortable comparing our estimations to those found in social media. If the reviewer has any papers in mind, which we've overlooked, we'll gladly consider a feasible comparison.

Did you expect that the ECHAM simulation reproduces the plume correctly? Why did you expect that?

Our objective was to investigate the performance of the aerosol microphysical module integrated into the ECHAM, model. As the model has already been successfully applied for the simulation of recent and past large volcanic eruptions (e.g., Niemeier et al., 2009, 2019; Toohey et al., 2016, 2019), we certainly expected to obtain physically reasonable values. However, it was unknown, if the shape of the plume and extinction values will be similar. Furthermore, sensitivity of the model to plume initialization and nudging was not investigated before.

How well did previous model studies simulate tropical UTLS injections? What are the error sources?

Please, refer to the comment above, ECHAM participated in the Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP) (Timmreck et al., 2018). One of the experiments deals with small to moderate eruption of the early 21st century and will provide useful information about general behavior of global aerosol models to simulate UTLS aerols. Additionally, the lifetime of aerosols for injections into the UTLS depends strongly on the representation of stratospheric dynamics in the model (Niemeier et a, 2020), mainly ω^* and sedimentation. The residual vertical velocity ω^* is driven by parameterized waves and a combination of larger-scale waves which are impacted by numerous aspects of the model such as horizontal and vertical resolution, diffusion parameterization and physics parameterizations. Sedimentation depends mainly on particle size which is strongly determined by aerosol microphysical processes and transport. Due to these non-linearities the error sources are manifold.

In how far does your aerosol uplift rate agree with the expected uplift rate from the water vapour tape recorder?

We did not perform any investigations of the uplift rate. An additional investigation of a water vapour dataset would clearly displace the focus of the paper.

You mention that it is probably faster due to additional heating, but no evidence is shown.

We agreed that mentioning it without an evidence is not ideal. We removed that sentence from the text.

Section 7.2: The radiative forcing is presented as method description and results. In my opinion this part rather belongs into the main part of the paper. Furthermore the analysis is not well thought through. The result is presented in a descriptive manner. I'd recommend to condense and get to the point.

The radiative forcing calculations are partially based on the extinction coefficients thus they should be presented after the presentation of the extinction. One of the main objectives of the radiative forcing section is a comparison of

calculations using different approaches from measured and modeled data. With that the radiative forcing section has to follow the model-observation comparison section. Thus, the only suitable place for this section is the end of the paper. We, however, agree that the assignment of this section to the "discussion" might be sub-optimal. This has been changed and hopefully improved during the restructuring of the paper, in line with suggestions from the reviewers.

Unfortunately the reviewer does not give us any information to understand what he/she means saying that the analysis is not well thought through and what kind of improvements he/she recommends and why. This gives us no chance to make any improvements with respect to this comment. Is anything wrong in a presenting of the results in a descriptive manner?

We believe that describing the methods of obtaining our result in a such crucial point as radiative forcing is answering the standards of reproducibility of the results. For legal reasons, we can't publish the code openly, but we believe that a description of how we obtained the results is the second best thing. The section is already quite short and we would need to drop some information if we condensed it. This will reduce the readability of the paper and may result in confusion. We do not see enough justification for shortening this section. In the section we prove the validity of the simple formula suggested by Hansen et al. for calculating the radiative forcing in a limited latitude range, estimated the radiative forcing from the Ambae eruption and demonstrated an overall agreement between the modelled and observed values of the radiative forcing.

l401-408: Here, merely the figure and the line colors are described. This rather belongs into the figure caption. Please explain the results and which conclusions can be drawn from the data shown in Fig. 6.

We fully agree, this paragraph describes the lines in the plot in more details than the figure caption does. We prefer to keep it for better clarity. The results are explained starting from the line 409 of the original manuscript and the conclusions are presented in the last paragraph of the section.

l418, 420: "between dashed and solid blue lines" and "panel a of Fig. 5": Please name the parameters represented by the colored lines or panels.

Concerning "between dashed and solid blue lines", this notation gives a clear reference to the plot and is much more readable than complicated names of the corresponding parameters. We prefer to keep it as it is. With respect to "panel a of Fig. 5" the text was changed to "in agreement with the maximum of the extinction coefficient seen in the panel (a) of Fig. 5.

l434: Why does your result differ significantly from Kloss et al., 2020? Please discuss.

The discussion requested by the reviewer is added to the manuscript text.

Technical:

Please write out all abbreviations on first use.

We went through the text, and followed the suggestion.

l14: injection estimates

Corrected.

l18: Which ECMWF reanalysis, ERA-interim or ERA5?

We used ERA5 for nudging. We changed the text to: ".. as well as through nudging to ECMWF ERA5 reanalysis data."

l24: ... climate system -is- now ...

Corrected.

l40: comma before which

Added.

l41: (UTLS); closing bracket missing

Corrected.

l59: used

Corrected.

l109: This results in the situation that the usual stratospheric aerosol extinction wavelength 750 nm, used by e.g. SCIAMACHY and OSIRIS (Rieger et al., 2018), is not suitable, as OMPS-LP measurements around this wavelength are affected by the O2-A absorption band.

Corrected.

l111,115,123: used

Corrected.

l117: for all the tangent altitudes $->$ for all tangent altitudes

Corrected.

l122: is $->$ was

Corrected.

l200: comma before because

Comma was added.

l226: "...below 20°S..." Do you mean south of 20°S?

l242: used

Corrected.

l246: do $->$ did

Corrected.

l261: -1E30 -> -1 $\times$ 10^{30}

Corrected.

l279: I assume this should be 6.1.2.

Yes, thank you. This assumption was correct. The numeration has changed though during the re-structure.

l274: ejected -> injected

Corrected.

306: "g/m22" remove 2

Corrected.

l317: 2x is -> was

Corrected.

References

Brühl, C., Schallock, J., Klingmüller, K., Robert, C., Bingen, C., Clarisse, L., Heckel, A., North, P., and Rieger, L.: Stratospheric aerosol radiative forcing simulated by the chemistry climate model EMAC using Aerosol CCI satellite data, Atmospheric Chemistry and

Physics, 18, 12 845–12 857, https://doi.org/10.5194/acp-18-12845-2018, https://www.atmos-chem-phys.net/18/12845/2018/, 2018.

Fromm, M., D. T. Lindsey, R. Servranckx, G. Yue, T. Trickl, R. Sica, P. Doucet, and S.Godin-Beekmann (2010), The untold story of pyrocumulonimbus, Bull. Am. Meteorol.Soc., 91, 1193– 1209, doi:10.1175/2010BAMS3004.1.

Kloss, C., Sellitto, P., Legras, B., Vernier, J.-P., Jégou, F., Venkat Ratnam, M., et al.(2020). Impact of the 2018 Ambae eruption on the global stratospheric aerosol layer and climate. Journal of Geophysical Research: Atmospheres, 125, e2020JD032410.https://doi.org/10.1029/2020JD032410

Vernier, J.-P., et al. (2011), Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, Geophys. Res. Lett., 38, L12807, doi:10.1029/2011GL047563.

Bourassa, A. E., D. A. Degenstein, B. J. Elash, and E. J. Llewellyn (2010), Evolution of the stratospheric aerosol enhancement following the eruptions of Okmok and Kasatochi: Odin-OSIRIS measurements, J. Geophys. Res., 115, D00L03, doi:10.1029/2009JD013274.

Bourassa, A. E., Rieger, L. A., Zawada,D. J., Khaykin, S., Thomason, L. W., and Degenstein, D. A. (2019). Satellite limb observations of unprecedented fores fire aerosol in the stratosphere. Journal of Geophysical Research: Atmospheres, 124, 9510–9519. https://doi.org/10.1029/2019JD030607.

Chen, Z., DeLand, M., and Bhartia, P. K.: A new algorithm for detecting cloud height using OMPS/LP measurements, Atmos. Meas. Tech., 9, 1239–1246, https://doi.org/10.5194/amt-9-1239-2016, 2016.

Krotkov, N. A., McLinden, C. A., Li, C., Lamsal, L. N., Celarier, E. A., Marchenko, S. V., Swartz, W. H., Bucsela, E. J., Joiner, J., Duncan, B. N., Boersma, K. F., Veefkind, J. P., Levelt, P. F., Fioletov, V. E., Dickerson, R. R., He, H., Lu, Z., and Streets, D. G.: Aura OMI observations of regional SO2 and NO2 pollution changes from 2005 to 2015, Atmos. Chem. Phys., 16, 4605–4629, https://doi.org/10.5194/acp-16-4605-2016, 2016.

Clyne, M., Lamarque, J.-F., Mills, M. J., Khodri, M., Ball, W., Bekki, S., Dhomse, S. S., Lebas, N., Mann, G., Marshall, L., Niemeier, U., Poulain, V., Robock, A., Rozanov, E., Schmidt, A., Stenke, A., Sukhodolov, T., Timmreck, C., Toohey, M., Tummon, F., Zanchettin, D., Zhu, Y., and Toon, O. B.: Model physics and chemistry causing intermodel disagreement within the VolMIP-Tambora Interactive Stratospheric Aerosol ensemble, Atmos. Chem. Phys., 21, 3317–3343, https://doi.org/10.5194/acp-21-3317-2021, 2021.

Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., Stein Zweers, D. C., Duncan, B. N., Streets, D. G., Eskes, H., van der A, R., McLinden, C., Fioletov, V., Carn, S., de Laat, J., DeLand, M., Marchenko, S., McPeters, R., Ziemke, J., Fu, D., Liu, X., Pickering, K., Apituley, A., González Abad, G., Arola, A., Boersma, F., Chan Miller, C., Chance, K., de Graaf, M., Hakkarainen, J., Hassinen, S., Ialongo, I., Kleipool, Q., Krotkov, N., Li, C., Lamsal, L., Newman, P., Nowlan, C., Suleiman, R., Tilstra, L. G., Torres, O., Wang, H., and Wargan, K.: The Ozone Monitoring Instrument: overview of 14 years in space, Atmos. Chem. Phys., 18, 5699–5745, https://doi.org/10.5194/acp-

18-5699-2018, 2018.

Muser, L. O., Hoshyaripour, G. A., Bruckert, J., Horváth, Á., Malinina, E., Wallis, S., Prata, F. J., Rozanov, A., von Savigny, C., Vogel, H., and Vogel, B.: Particle aging and aerosol–radiation interaction affect volcanic plume dispersion: evidence from the Raikoke 2019 eruption, Atmos. Chem. Phys., 20, 15015–15036, https://doi.org/10.5194/acp-20-15015-2020, 2020.

Timmreck, C., Mann, G. W., Aquila, V., Hommel, R., Lee, L. A., Schmidt, A., Brühl, C., Carn, S., Chin, M., Dhomse, S. S., Diehl, T., English, J. M., Mills, M. J., Neely, R., Sheng, J., Toohey, M., and Weisenstein, D.: The Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP): motivation and experimental design, Geosci. Model Dev., 11, 2581–2608, https://doi.org/10.5194/gmd-11-2581-2018, 2018.

Theys, N., De Smedt, I., Yu, H., Danckaert, T., van Gent, J., Hörmann, C., Wagner, T., Hedelt, P., Bauer, H., Romahn, F., Pedergnana, M., Loyola, D., and Van Roozendael, M.: Sulfur dioxide retrievals from TROPOMI onboard Sentinel-5 Precursor: algorithm theoretical basis, Atmos. Meas. Tech., 10, 119–153, https://doi.org/10.5194/amt-10-119-2017, 2017.

Vernier, J.-P., Fairlie, T. D., Deshler, T., Natarajan, M., Knepp, T., Foster, K., Wienhold, F. G., Bedka, K. M., Thomason, L., and Trepte, C.: In situ and space-based observations of the Kelud volcanic plume: The persistence of ash in the lower stratosphere, J. Geophys. Res.-Atmos., 121, 11104–11118, https://doi.org/10.1002/2016JD025344, 2016.

Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E. P., Hegerl, G., Robock, A., Pausata, F. S. R., Ball, W. T., Bauer, S. E., Bekki, S., Dhomse, S. S., LeGrande, A. N., Mann, G. W., Marshall, L., Mills, M., Marchand, M., Niemeier, U., Poulain, V., Rozanov, E., Rubino, A., Stenke, A., Tsigaridis, K., and Tummon, F.: The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): experimental design and forcing input data for CMIP6, Geosci. Model Dev., 9, 2701–2719, https://doi.org/10.5194/gmd-9-2701-2016, 2016.

Short comment: Pasquale Sellitto

Dear Authors,
thank you for your paper, which I found very interesting and an enjoyable reading.

Dear Pasquale, thank you very much for a careful reading of the paper and providing your comments.

I would strongly suggest to compare your aerosol extinction (and AOD) as well as your radiative forcing (RF) estimations with those previously published by Kloss et al., JGR,2020 (that you cite in your manuscript).

Please bear in mind that our paper is focused on a comparison of the measurement and modelling results and a comparison of the aerosol extinction with Kloss et al. (2020) would displace the focus and make the paper more lengthy

and difficult to read. We would like to highlight that the only aerosol extinction result, which could be compared without making additional assumptions needed to convert to other wavelengths, is the averaged SAGE III/ISS profile at 869 nm, which is shown in Fig. 8, and the corresponding AOT. As our paper already presents a comparison with SAGE III/ISS data for the period of interest (Fig. 1 of our paper) and the agreement in tropics is found to be better than 25% we do not think a comparison of the averaged extinction profiles with those from Kloss et al. (2020) will bring any additional value to our research topic.

For the RF estimations, in particular, Kloss et al. have used a detailed radiative transfer modelling, which is expected to bring more precise estimations than the rough parameterisation of your Eq. 1.

One of the topics, discussed in our paper, is how well the simple approximation provided by Eq. 1 performs in comparison with the internal flux calculation of the ECHAM model. The latter is expected to be much more precise. The question, what is more precise, a radiative module of a GCM or a standard radiative transfer model, is open for now and cannot be answered in the framework of our study. Radiative transfer calculations in a GCM are based on strong approximations but the employed approach is optimized for broad-band flux calculations. A standard radiative transfer model is optimized for precise calculations of spectrally resolved radiances but is not necessarily the best choice if a small difference of two large values (fluxes) needs to be calculated as it is done in the case of the radiative forcing. Furthermore, the influence of the vertical and angular grids, Fourier harmonics and possibly atmospheric sphericity needs to be carefully investigated before making any conclusions.

Please also specify that the RF of Eq. 1 is the top of the atmosphere RF or otherwise clarify at which altitude it is estimated.

The RF is calculated for the top of the atmosphere. We will add this information to the revised manuscript.

Kloss et al. included also several hypotheses on the spectral variability of the extinction, the absorption properties of the particles and the angular distribution of the scattered radiation, and obtained quite different results if compared to yours. This should certainly be discussed in your manuscript. Please add such comparisons to your revised manuscript, that would be a great added value to your interesting work.

ECHAM features a built-in aerosol module which models aerosol microphysical processes to calculate the aerosol composition and size distribution. This information is used then to calculate fluxes. It is not obvious how to compare an aerosol parameterization as represented by the model with assumptions made by Kloss et al. especially with respect to an outdated parameterization of the aerosol scattering with an asymmetry parameter. On the contrary, the values of the radiative forcings can be compared and the reasons for their disagreement will be discussed in the revised manuscript. In our opinion, the main reason for disagreement is the way the non-Ambae signal has been subtracted. By subtracting the radiative forcing just before the second eruption we ensure that only the Ambae contribution is accounted for. On the contrary, Kloss et

al. use some "background aerosol" scenario, which is not clearly described in the paper, to remove other contributions than that of Ambae. We are quite sure that the state of the stratosphere just before the Ambae eruption was far from background (in the common sense of this term), e.g. because of a residual aerosol load from Canadian wildfires 2017. Another issue can be related to the definition of the tropopause. Although it is not stated clearly, from the remark "(full profiles, for the whole stratosphere)" in Kloss et al., we guess that only the stratospheric part of the aerosol profile is used to calculate the radiative forcing. This is the same approach as we used. If our guess is wrong and Kloss et al. use the tropospheric part of the profiles as well, this might be a reason for the disagreement. Otherwise, there might still be a difference in a definition of the tropopause. Unfortunately, Kloss et al. do not state how their tropopause was defined.

If the Editor thinks it useful, I could review the revised version of the manuscript, in particular for what concerns this comparison and the RF estimations themselves.

If you are interested in understanding possible reasons for the disagreement and evaluation of different methods, the only possible way to do it, would be a dedicated joint investigation which is far beyond the scope of our study. We see no possibility to draw any well justified conclusions based only on the information published by Kloss et al. (2020). If you still have an interest in a joint study, please contact the corresponding author (Alexei Rozanov) of the manuscript.

My best regards,
Pasquale Sellitto