We thank the reviewers for constructive feedback that has improved our manuscript. The reviewer comments are given below in black font followed by our answers in this blue color.

General points:

1. At the end of the abstract, a “gridded high vertical resolution SO_2 inventory that can be used in Earth system models” is mentioned. The retrieval of the SO_2 height distribution in AIRS swaths is well described, but I’m not sure I would classify that as “gridded inventory”. Of course, for global Earth system models, SO_2 data on a regular horizontal/vertical/temporal grid would be of great value. If your data product is anything like such a larger gridded inventory, then please do describe this in more detail (on page 10, you describe gridding the FLEXPART output on a 0.5 x 0.5 degree grid, but that is prior to linking it to the AIRS information, so it is not your final “product”, or is it?).

We agree with the reviewer. We have compiled a 3D-gridded products and added a paragraph on the 3D-dataset in the results section. “…The dataset has 1° latitude × 1° longitude horizontal resolution and comes in the three different versions depending on which vertical coordinates that is preferred (potential temperature, geometric altitude, and pressure). The vertical resolutions are 1 K for the potential temperature grid, 200 m for geometric altitude, and for the pressure grid the pressure levels correspond to geometric altitude steps of 200 m. The dataset is available as a supplementary file…”

2. In the same context, the simpler (compared to a “gridded inventory”) vertical information for volcanic plumes/clouds/sub-clouds may still be very useful for modellers. But I wonder if hinging them to the AIRS swaths in space and time is the most useful thing that can be done. If you assume that the vertical distribution of the aerosols observed by CALIOP is representative for the SO_2 distribution, then why not calculate everything right back to the volcano? Vertically resolved emission plumes could then be “injected” into models of any scale.

We agree that it could be useful to use only few vertical profiles (one per eruption) at the geographical location and time of eruptions. However, the aim of our study is to provide SO_2 profiles by combining the different strengths of the different satellite instruments. Our data-set can be implemented directly into models, and would likely be at least as good as injection profiles from the volcanoes site, maybe better: Starting with a single (or few) horizontal grid point(s), small errors in simulated transport paths in the first few days may lead to larger errors than starting out with a SO_2 distribution already spread over a larger area. This is particularly useful for Earth system models with coarser resolution that may have limited ability to capture the initial transport.

3. There has been a rather comprehensive study on the Sarychev plume and its dispersion by Wu et al. (2017), and I was rather surprised to see no mention of it in your paper! I suggest you look carefully at this previous work that also combines satellite observations and trajectory calculations. It should probably be mentioned in your introduction, and a comparison of the two studies (What was done similar? What different? Are the results in agreement?) in your discussion is clearly warranted. Note that while I have been working on this review, a short comment was uploaded by the authors of the Wu et al. (2017) paper. This comment contains some excellent ideas and suggestions as to how connections between the two studies could be made, and I strongly recommend you to carefully look at this short comment and follow the recommendations.

We were not aware of the Wu et al. paper when writing the manuscript, and are relieved that the authors contacted us at such an early stage. We have contacted the authors and added their data to Fig. 7 (now Fig. 8). We added their study in the introductions section and added a reference to their study.

4. According to your method description, you use ERA-5 reanalysis data to drive the FLEXPART trajectory model and MERRA-2 to determine tropopause heights and potential temperature levels of the CALIOP data (Page 7, lines 22 – 25). Would it not be more consistent to use one single reanalysis data set for both?

We agree that it is generally preferable to use the same reanalysis data. However, FLEXPART uses the ERA-5 data, whereas the CALIOP hdf-files contain meteorological data from MERRA-2 simulated to the specific locations and time of the sensor's observations.

5. Can you give more detail on how horizontal transport and transforms between different vertical coordinates is realized? On page 9, lines 6 – 10, you explain that potential temperature "is stored" for the released tracer particles. On page 10, lines 24 – 26, you state that you "did not use the FLEXPART output vertical coordinates" and "set a single height interval between 0 and 50 km in the output grid specification". I find this confusing. Do you mean that you transport the parcels from the CALIOP SWATH converting geometric altitude to potential temperature in the beginning and then convert the same potential temperature value back to geometric altitude at the end? And when you say "locations" on page 9, line 10, does that include time in that case? Yes, it includes time. We see that this may be confusing. Thank you for pointing this out. We have made adjustments in the manuscript to clarify how this procedure works (at the end of section 2.2). The geometric altitude is first transformed to potential temperature, computational particles are then transported, and finally we transform back to geometric altitude. (The first step is to convert geometric altitudes to potential temperatures. Each computational particle gets a potential temperature value that follows the particle during its transport. After transport, the particles are put on a potential temperature grid to sum up the amount of particles in each potential temperature bin.).

If the potential temperature of an air parcel is indeed not allowed to change during transport in your simulations, then the method would obviously not work in the tropics, where cross-isentropic vertical transport is common. Also, self-lofting in fresh volcanic plumes would not be accounted for, correct? Some clarifications and a more extensive discussion of this would help.

It is true that our method cannot capture diabatic transport, but this should be a minor issue due to the short time span needed for the analysis. It is important to use potential temperature to follow the aerosol transport in the stratosphere. Dense aerosol clouds of supermicron particles can lead to radiative heating and self-lofting, but this effect should have little impact on the smaller particles from the Sarychev eruptions. We agree that it would be difficult to use this approach for a situation where we have rapid radiative heating of a dense aerosol layer with supermicron particles like dense wildfire smoke. This may be the case also for Mt Pinatubo's and larger eruptions' volcanic clouds. For the present eruption, similar-sized, and smaller eruptions, the potential temperature should be contained during the short time-frame in our study (few days backwards and forwards). It is worth noting that Fairlie et al., (2014) found a self-lofting of no more than 0.3 K/day for aerosol from the

June 12, 2011 eruption of Nabro. Nabro injected similar amounts of SO2 as Sarychev did. Hence, self-lofting should not have any major impact on our SO2 profiles.

We added a statement on self-lofting and its impact on our data in Section 3.1.

6. I do wonder to what extent the issues and uncertainties I mentioned under points 4 and 5 above could factor in producing the bimodal distribution in the sum of your study that is shown in Figure 7, and also on the height of your SO2 plume being somewhat higher than the earlier studies you mention. First, the relationship between potential temperature and geometric altitude is probably not the same at the different locations (I guess the profiles from all the earlier studies are close to the volcano in space and time?), so when you strictly transport along isentropes, it may be better to do this comparison on a potential temperature vertical coordinate. Second, self-lofting could explain at least part of the difference in altitude and should be mentioned.

The vertical distribution in our figure is in agreement with CALIOP observations both in July and few months after the eruptions of Sarychev (Fig. 8 in Friberg et al. 2018). The bimodal structure has to do with multiple eruptions from the volcano over the course of a few days. This is indicated in Fig. 6 in the manuscript where we see different vertical profiles at different locations. The Eastern parts are located at ~3 km (40 K) lower altitudes than the Western. With the numbers given by Fairlie et al. (2014), the lofting in our study period (9 days) would be 2.7 K (~160 m). Note that this would be the maximum possible lofting according to their study, i.e. much smaller than the height difference between the different sub-clouds. Given these numbers, self-lofting cannot explain the bimodality in Fig. 7. More details regarding self-lofting are discussed in our answer to comment #5 above.

The potential temperature was converted back to geometric altitude at the locations of the AIRS pixels. Hence, there should be no mismatch from our approach. We agree that it would be preferable to compare all studies on potential temperature coordinates, but most studies only report geometric altitudes or pressure levels. Thus, we preferred to transform our dataset to the individual AIRS swaths' locations instead of assuming a climatological transform function for the published data that we compared our data to.

7. As you state on page 3, lines 8/9, an important assumption for your method to work is the horizontal and vertical co-location of volcanic SO2 and aerosol following the eruption. While this assumption is probably reasonable for a wide range of eruptions and conditions, I still suggest a more detailed discussion to further support this assumption: what do we know from the literature with respect to the SO2/aerosol co-location? Are there conditions, under which the assumption may not hold, and on what time scales will it hold? How about ash-rich eruptions, where particle sedimentation rates may be higher (you mention this at the end of page 6)? This may all not be critical for your particular case looking at rather short time scales not so long after the eruption for a non-tropical eruption. But as you describe a new method that should hopefully be applicable to other cases, such caveats and limitations should be made more clear.

We agree that it is important to consider possible ash signals. CALIOP retrieves a depolarization ratio that we use as a means of checking the impact of ash. We investigated the homogeneity of the depolarization ratios' vertical distributions and find no evidence of strong sedimentation during the studied period (section 2.2). The fact that the vertical distribution in the present study matches CALIOP observations in weeks and months later is evidence for that ash has little impact on our retrieved SO2 profiles. We agree with the reviewer that it is valuable to clarify that it may differ depending on the eruption and added text to the manuscript clarifying that this may not always be the case, but that it worked fine for Sarychev (Section 2.2).

Specific and minor points:

Page 2, line 15: The Watts (2000) reference for OCS sources and sinks is in the middle of the timeline of possible references, i.e. it is neither the first one nor is it up to date. I suggest to cite Kremser et al. (2016, already in your reference list), which contains a fairly recent OCS budget together with a discussion of source and sink processes. Thank you for pointing this out. We now refer to Kremser instead of Watts.

Page 9, lines 1 – 5: I'm not sure that I have fully understood the scaling of the number of FLEXPART particles. Are you saying that you are releasing x FLEXPART particles per one CALIOP pixel, and x scales with β_{aer} of that pixel? Yes, the released number of particles scales with the strength of the CALIOP signal (light backscattering). This enable us to run CALIOP swaths with both strong and weak backscattering signals. We changed the text to clarify (Sections 3.2 and 3.3).

Page 12, lines 2 – 3: As the reason for the lack of trajectories matching some of the AIRS SO_2 clouds, you mentioned that these SO_2 detections could be located below the tropopause, while only the stratospheric aerosol scattering observed by CALIOP is "transported". Have you tried testing this by, e.g., using TP - 2km instead of TP as vertical threshold? We have checked the occurrence of volcanic aerosol in the vicinity of the tropopause. It is difficult to separate ice-clouds from volcanic aerosol in the troposphere, and more so the further away from the tropopause. We therefore cannot include the troposphere in our analysis.