Review of manuscript “Volcanic SO_2 Layer Height by TROPOMI/S5P; validation against IASI/MetOp and CALIOP/CALIPSO observations.” by Koukouli et al.

General comments:

This paper compares volcanic SO_2 Layer Height (LH) retrievals from TROPOMI/S5P with the retrievals from the IASI/MetOp and CALIOP/CALIPSO instruments for several recent volcanic eruptions. Based on the analysis presented, the authors find that the average S5P SO_2 LH product compare well with the LH estimates from the IASI and CALIOP/CALIPSO instruments. Furthermore, the authors find that assimilating the S5P SO_2 LH product into the CAMS model improves the forecast of the 2019 Raikoke volcanic plume.

Overall, the paper provides a useful evaluation of the SP5 SO_2 LH product and the topic of this paper will be of interest to the readers of this journal. The results show that these new S5P plume height estimates are of equal quality as other similar products and is a useful addition to the existing range of volcanic plume height estimation techniques/products available in the literature. The paper is well structured, and the quality of the figures and tables is good.

However, I do have some concerns regarding the discussion of the results. In general, I miss an in-depth discussion of the implications of the results and the potential additional new information we can obtain from considering this new product. The paper presents the results of the comparison of S5P with the other satellite products, but not much is presented in terms of discussion of their findings and the differences found between the different retrievals. (e.g. What are the main limitations of the S5P LH product? What are the main advantages of using the S5P LH product with respect to the IASI SO_2 LH estimates? Why is the distribution in figure 3b for TROPOMI over a much wider range compared to the IASI estimates?)

This study also mainly focusses on the mean LH values and how they compare between the satellite products. However, for most volcanic eruptions, the emissions take place over a large range of altitudes (as is also evident by the large standard deviations reported in the manuscript, e.g. tables 2 & 3). The comparison of the distribution of the plume altitudes is therefore potentially equally relevant. Therefore, I think the paper would benefit from a more detailed discussion of the differences between the LH distributions (e.g. as shown in figure 3) and the corresponding statistics.

Furthermore, I found a lot of small inconsistencies between the values reported in the text and the tables. The level of significance for the reported values of the same quantities are not consistent (e.g. average SO_2 LH in tables 2 and 3, versus values in L.274-277). The reported values should be made consistent throughout the manuscript.

Finally, section 4.3 shows very similar results as presented by Inness et al. 2021. The difference is the use of a later version of the TROPOMI SO2 LH product (see L.445: “The assimilation of the S5P SO2LH data was based on a previous version of the dataset, v3.1, and not the final one, v4.0 presented in this work”). However, no information is given what the differences are between the two versions. The general conclusions presented here are

very similar as presented by Inness et al. 2021. I think a more detailed discussion of the differences between the two studies should be included, as otherwise this section provides very limited new information to the scientific community.

I would recommend major revisions before the paper can be considered for publication.

Specific comments:

Title: 'validation' I am not sure this is the right term. As there are a lot of uncertainties also in the IASI/MetOP and CALIOP/CALIPSO retrieval algorithms, we can't be sure they represent the true values either. Instead, 'evaluation' or 'comparison with' might be better terms to use.

L.28: 'satisfactory' How would you define if a comparison is satisfactory? It depends on the application for which you want to use the SO_2 LH product and is case specific.

L.31: I think there is a comma missing after '1.5±2km'.

L.42: The used reference (ICAO, 2012) doesn't mention SO_2 clouds and is mainly focussed on the risks posed by ash clouds. I think a different reference should be used here, for example: https://www.icao.int/airnavigation/METP/MOGVA Reference Documents/IAVW Roadmap.pdf (page 12)

L.59: 'validation' -> 'evaluation' or 'comparison'. See comment about the title.

L.261: 'Figure S1, … associated with loads of less than ~20 DU.' This is not clear from figure S1, as the colour scale range is starting at 20 DU for all panels. The only difference figure S1 shows is that the IASI estimates have a larger region where the SO_2 load is >20 DU. Why is the extent of the 'dense' plume in the IASI retrievals so much bigger than what is retrieved by TROPOMI?

L.267: 'well placed in height'. Is this correct? Figure 2 shows that the IASI AOPP product for the integrated SO_2 mass peaks 1-2 km higher than the estimates from TROPOMI and IASI ULB/LATMOS. As we are near the tropopause height, this change in peak altitude can mean the difference between most of the plume reaching the stratosphere or not. What could be the cause this difference?

L.268: If I understand correctly, when all pixels are excluded in IASI AOPP for a single grid box, the grid box is excluded from the presented comparison. Therefore, as each of the considered gridded data points is a collection of multiple pixels, could the exclusion of several pixels within a grid box explain the observed differences? Assuming that an average is calculated for each grid box, the difference seen would indicate that only very high concentration pixels are excluded by the IASI AOPP quality control. Is this the case, and if so, why would the IASI product have this bias?

L.275: ‘the mean S5P SO2 LH is reported at 10±3 km’. This is not consistent with the values presented in figure 3 and table 2. According to the legend in figure 3a and table 2, the estimate is 11±3 km (using the correct rounding).

L.276: ‘IASI AOPP placing the plume at 10±1 km’. How does this follow from Figure 3? I think the values for the two IASI products in the text are swapped, as it also does not correspond to the values reported in tables 2 and 3.

L.313: What are the values 2.5 and 4 km based on? In tables 2 & 3 neither of these values are present, so I am not sure I understand how these values are calculated.

L.324: The results presented in this work are heavily biased towards 1 eruption (Raikoke). How does this impact the statistics presented? If we exclude the other eruptions, what would the correlation coefficient be?

Fig.4: What are the uncertainty ranges of the slope and intercept calculated for the best linear fit? Is this represented by the blue shading in the figure? Some more explanation is needed in the caption on the blue shading and uncertainty ranges should be reported in the manuscript.

Fig.5: Please refer to figure 6 in the caption for the path of the CALIPSO satellite, as it made it easier for me to interpret figure 5.

Fig.6: Are the 2 colour scales different? I found it very confusing to have two very similar colour bars. If they are the same, I think it is better to use the same colour bar for both retrievals and have a double colour bar title instead.

Fig.7: In the right panel, what is the uncertainty in the ‘best’ slope calculated?

Fig. 8: The correlation between individual TROPOMI and CALIPSO pixels seems to be low when considering all the points in figure 8b. However, when considering the daily average values in figure 10, the comparison is much better. Is this because the points in fig 8b are clustered by day, therefore giving a better correlation for each of the individual overpasses? Might be useful to use 7 different colours in figure 8b to indicate the different days.

L.382: Related to the previous point, is there an impact of the aging plume on the results found? For example, can we expect the differences between CALIPSO and TROPOMI to be larger for older plumes due to the different dispersion of ash and SO_2?

L.446: What are the main differences between version 3.1 and 4.0? Do you expect there to be a big difference in skill between the two versions? Comparing the reported CAMS forecast bias of -1.5±2.5 km (L.475) in this manuscript with the value reported by Inness et al 2021 (0.4±2.2 km), it seems that the latest version 4.0 is less accurate. I think some discussion of this fact and potential reasons/implications should be included.

L.470: Are these LH values correct? Based on the results in tables 2 and 3, the values should be 11.4±2.5 km for IASI AOPP and 10.8±3.5 for S5P. Also, I am not sure I understand where

the 10.5 km comes from, as I can't find it anywhere in the results section (I think it should be 10.8 km based on table 2). Please check that all the values are correct.

L.471: Why are the results for the IASI ULB/LATMOS SO_2 LHs presented at a smaller accuracy in the conclusion section compared to table 3 (i.e. 0±3 km instead of -0.2±2.8 km)? Please make sure all the values in the text are consistent with the values presented in the tables/figures.

L.473-475: Some of the reported values for the Taal eruption are inconsistent with the values presented in table 2 and 3.

L.477: Different accuracy of the values than what is presented in table 3.

L.478: 'both sensors report high plume altitudes, at ~15km with both IASI/AOPP and ULB/LATMOS standard deviation at ~1km'. This is not consistent with the results section. Based on table 2, the LH for IASI AOPP is 13.5 km with a standard deviation of 3.4 km. Also, the S5P standard deviation is different in the two tables (2.5 km and 3.9 km) compared to the 4 km reported here.

L.491: "(low)" what does this word refer to? I think you mean that both showed low altitude plumes, but please add some additional text to clarify.

L.495: "quite closely". This is not a scientific term and should be avoided.

Technical corrections/suggestions:

L.25: 3km -> 3 km There are several other values in the abstract and the rest of the text where a space is missing between the value and the unit. I have tried to highlight most of them here, but please check carefully throughout the manuscript.

L.50: "volcanic processes assists in" missing comma

L.120: 20DU -> 20 DU. This is the first mention of DU, so I think it should be spelled out here.

L.127: D.U. -> DU. Throughout the manuscript both 'DU' and 'D.U.' are used. Please check and only use one consistent abbreviation.

L.157: 25km -> 25 km

L.194: 100m -> 100 m

L.225: Missing bracket after TROPOMI

L.242: '*including* both ascending'

L.264: 1km -> 1 km & 20km -> 20 km

L.274-275: PH -> LH ?

L.284: high -> thickness?

L.287: 15km -> 15 km

L.293: location -> altitude?

L.306: 1km -> 1 km & 14-15km -> 14-15 km

L.312: 1km -> 1 km

L313: 4km -> 4 km

L.373: 2km -> 2 km

L.374: that -> than

L.388: omit ‘a’ & 17km.. -> 17 km.

L.390: axis.Its -> axis. Its

L.393: Why is Figure 9 in bold font?

L.421: omit ‘nearly’

L.423: 1km -> 1 km

Figure 11: Please expand the caption by explaining the BLexp (no assimilation) and LHexp (assimilation of TROPOMI data) terms.

L.469-498: a lot of places where a space between the value and unit is missing.

L.492: 7.km -> 7 km

L.482: 0.72 -> 0.73. Based on L.324 I think this should be 0.73