Review Comment on amt-2022-148
Anonymous Referee #2


The study "Improved retrieval of SO2 plume height from TROPOMI using an iterative Covariance-Based Retrieval Algorithm" by Theys et al. introduces an optimal estimation retrieval algorithm to retrieve SO2 column and layer height of volcanic injections from nadir measurements between 309 and 329 nm by TROPOMI. First, the retrieval is described and its theoretical uncertainties are assessed. Then examples of the retrieval for various volcanic plumes are presented. The layer heights are verified by comparison with IASI measurements and reconstructed plume heights from a backward trajectory approach. In general I found this paper well structured and well written. Hence I recommend this paper for publication after addressing the minor comments below.

**General comments:**
This paper contains many abbreviations, which often hamper the reading flow (e.g. The LER albedo is retrieved by matching the measured mean radiance to a LUT of radiances (generated in parallel to the SO2 OD LUT), and which depends on SZA, VZA, RAA, surface height and albedo, with the same grid definition as in Table 1.). I suggest using less abbreviations, i.e. write out LH, LER, LUT, OD, VCD.

Please note, usually it is called a volcanic "injection" if a volcanic eruption reaches the upper troposphere and stratosphere. Volcanic "emission" is rather used for low altitude degassing. Please consider replacing "emit, emission,..." by "inject, injection,..." (e.g. p2ll6,9,11).

For the setup of the LUTs the US standard atmosphere was used. This is rather representative of midlatitude conditions. Could this have a (negative) impact on the retrievals of plume heights close to the tropical tropopause? Did you perform sensitivity tests for tropical, midlatitude, and polar atmospheric conditions?
Specific comments:
p2l4: Please specify what "difficult observation conditions" means.

p6l6: Please explain "Lambertian Equivalent Reflectivity" or add a reference.

p6l7: "More details are given below." Please specify details on what are given? Simulation setup? Simulation output parameters?

p6l13: Please add one sentence describing the "so-called solar I\textsubscript{0} effect".

p6l14-15: It is not clear to me if 450 refers to the across-track pixel position or to the ISRF parameters.

p6l30: Why does the LER approach work only well in principle? How is it in practice? Please provide a reference.

p8l21: Is there a default value for VCD\textsubscript{0}? Which one?

p9l9: Is minimum height + 1 km = surface height 1 km?

p9l10: Is maximum height - 1 km = 24 km?

p9l18: How much is 300 pixels roughly in km?

p9l20-22: Is there an upper limit in terms of distance to the central pixel?

p10: How did you predefine the Jacobians and the covariance matrix?

Figure 2: Is this the "sweet spot" configuration where the layer height error is smallest for low DU? For which surface type (i.e. ocean, land, ice) is an albedo of 5% representative? In the text you mention that you tested the impact of the albedo and SZA. What about the impact of surface height, ozone, RAA and VZA?
What are S-5P FP_ILM results?

How do you know from TROPOMI data that the plume had multiple layers?

Fig. 3: Could you please add contours to DU>20 to all panels or show contours for DU>5 to the DOAS panel? I see the point that you want to compare the layer heights, but I find the different shapes of the plumes confusing. If I understood correctly the DOAS method can retrieve lower DU, but has a larger uncertainty for <20 DU. Also, please add the orbit footprint to make the gap between two orbits better visible, especially in the region around the volcano.

Fig. 4: Please add the orbit footprint here too. Also I think it makes sense to cut Fig b and d at 20 km, as the color bar in a and b is also cut at 20 km.

Could it be that the larger difference to CALIOP aerosol height is due to the fact that for CALIOP the aerosol layer top altitude is compared to the TROPOMI layer height that is more representative of the center of the SO₂ layer?

Fig. 5: See general comment on using only the US standard atmosphere instead of representative profiles for the tropics and midlatitudes. How well do the TROPOMI and IASI results compare to CALIPSO aerosol heights?

Please specify "difficult conditions".

The method of using backward trajectories from satellite observations of plumes back to the volcano has been also used in other studies, e.g. Wu et al., 2017, Cai et al. 2022.

Does PlumeTraj only rely on position and time, or does it also consider the DU at each pixel to weight the reconstructed heights (e.g. high weight for large DU, low weight for small DU)?

Could this difference in plume height at the plume edges be also due to low DUs at the edges? How can you tell there is no underlying low altitude plume? Which altitude range is considered by PlumeTraj? Would this "edge-effect" still be there if you used e.g. 10 DU as a selection criterion for the layer height retrieval?
Fig. 6: According to Fig. 7 Fig. 6 shows the best case. What does a bad case, e.g. 3 March or 4 March look like? Is there also an ‘‘edge-effect’’ visible?

p22l12: How old is ‘‘old’’? Days, weeks?

Fig. 8: The figure quality needs to be improved. I don't understand why there are lines for TROPOMI.

p25l8-10: How were the deviations calculated? On a daily basis? There are lines for TROPOMI and points for IASI data. Sometimes there are data points for IASI, but not TROPOMI data.

Technical comments:
p2l4: expect -> except

p3l27: remove ‘‘obviously’’

p5l16: $x_i$ -> $x_i$

p6l28: aerosols layer -> aerosol layers

p7l2: aerosols -> aerosol particles

p7l5: Please introduce the abbreviation OCRA/ROCINN CRB.

p7l9: Please introduce the abbreviation RAA.

p13l26: Please introduce the abbreviation CrIS.

p13l31: sensors -> measurements
p17l20: observed a -> observed an

Fig. 5 caption: 2018 -> 2019

p22l29: Example -> Examples

References:
Wu et al., 2017; https://doi.org/10.5194/acp-17-13439-2017
Cai et al. 2022; https://doi.org/10.5194/acp-22-6787-2022