

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-219-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-219

Nina Iren Kristiansen (Referee)

Referee comment on "Evaluating the assimilation of S5P/TROPOMI near real-time SO₂ columns and layer height data into the CAMS integrated forecasting system (CY47R1), based on a case study of the 2019 Raikoke eruption" by Antje Inness et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-219-RC2>, 2021

Review of "The CAMS volcanic forecasting system utilizing near-real time data assimilation of S5P/TROPOMI SO₂ retrievals" by Antje Inness et al.

Review by Nina Kristiansen (UK Met Office)

General comments

This paper presents the CAMS assimilation of volcanic SO₂ satellite data, and in particular improvements made to the system by the use of layer height information retrieved from satellite, which show to improve the SO₂ forecasts. The paper is interesting and presents both improvements to and current challenges with the system. The topic of the paper is highly relevant as it addresses a method which can be used to fuse models and observations and targets a particular application to volcanic clouds. The paper is well written and highly suited for publication; however, I would like the below comments to first be addressed.

I miss some discussion around the applied/assumed thickness of the SO₂ plume and if/how this might affect the results. See specific comment on L260.

I am concerned that the model simulations do not directly consider the vertical averaging kernel information from the SO₂ retrieval. See specific comment on L262.

I miss some further details on the TROPOMI SO₂ retrieval. Other TROPOMI SO₂ total column retrievals are interlinked with assumptions on the SO₂ plume altitude and often different products are available based on different a priori plume altitudes (e.g., the Copernicus SP5 products). L122 mention prior SO₂ vertical profile shapes but this is not mentioned again or discussed any further for the DLR TROPOMI retrieval (only for IASI on L186). Please elaborate further on which prior profiles are used in the DLR TROPOMI retrievals (both NRT and LH) and if these vary, and also how/if this affects the retrieval of the layer height.

Can you provide some indications as to how much more expensive (in terms of run time) the model runs are with the higher spectral resolutions used? For example, it would be very interesting to know the difference in run time for each of the experiments in Table 3.

Specific comments

L30 - the last sentence of the abstract: It would be good to include here something about the increase in skill time scales by including the LH information. I would also include a couple more key results here; that including LH information leads to higher modelled TCSO₂ values in better agreement with the satellite observations, but that plume area and burden are overestimated also when including LH data and that the reason for this overestimation is explored.

L40: "SO₂ in the aircraft cabin is the biggest issue leading to respiratory problems for passengers and crew". Respiratory problems related to SO₂ depend on the SO₂ concentrations/dose and air quality standards for SO₂ exist. Potential problems also depend on people's underlying health problems like asthma. It doesn't therefore always lead to respiratory problems as this sentence seem to indicate.

L80: You use the different terms 'injection height' / 'plume height' / 'layer height' but not consistently and the difference between them (if any) is not explained. Personally, I'd use injection height as above the volcano and plume/layer height for the cloud altitude away from the vent, but it might be best to keep to as few terms as possible throughout the paper.

L135-147 (section 2.2): I miss some details on how the retrieval of the LH is done and what it relies on besides the exact wavelength ranges used. In which cases does it work well and which not (see related comment on L416). Why does it not work well below 20 DU? Also, what does this LH mean physically? You later use the height of the modelled maximum concentration as the model equivalent, would be good to comment on this here to justify that that is appropriate.

L207/section 3.2: The reader needs to know quite a bit about 4DVar assimilation systems

to follow this section. It would be good to expand a little in particular on those aspects which you later explore in more detail: background error covariance matrix and the minimisations. Also, observations errors are not mentioned at all, how are errors in the observations taken into account?

L260: "calculate the SO₂ column not between the surface and the top of the atmosphere, but between the pressure values that correspond to the bottom and the top of the retrieved volcanic SO₂ layer. The depth of this layer is currently set in the FP_ILM retrieval as 2 km, which corresponds to the uncertainty of the retrieved layer height." I am a little confused about this. Does it mean you use a fixed plume thickness of 2 km to calculate the modelled total columns, i.e., that you only calculate the SO₂ column between the bottom of the plume (retrieved LH - 2 km) and the retrieved LH? What if there is a much thicker plume say several km thick, then the calculation of the SO₂ column loading will miss a large fraction of the SO₂ in the vertical by only summing only over the LH-2km depth.

L262: "This approach mimics the procedure of using averaging kernels with box profiles given for the SO₂ layer.". I don't understand how this mimic the use of averaging kernels because if applying an averaging kernel sensitivity, the model data would be multiplied with a different sensitivity/ averaging kernel (AK) value at different vertical levels. Please elaborate. Ideally the satellites vertical AK profiles should be applied to the model data prior to any comparisons to the satellite data - this AK profile can vary from one satellite pixel to the next.

L278: "The 'dip' in the TROPOMI SO₂ burden after the initial peak is an artefact that results from missing observations in the TROPOMI NRT data." This 'dip' is not seen in the equivalent time series shown in the de Leeuw paper (their Fig 11) which also show TROPOMI data (different retrieval method). What is the cause of these 'missing observations'?

L300 / Table 3: From the order of the experiments given in the table I expected first the difference between the BLexp and LHexp to be discussed, however the LH50/100/250 cases are first discussed. Perhaps guide the reader at the start of the section to say which experiments are compared first and why. Similarly, would be good there to guide the reader to say that the BLexp and LHexp will be further explored later to assess the skill timescales to see if using a more realistic height rather than the default 5 km will improve the forecasts - a key point and question for the paper.

L415: "TROPOMI NRT lower detection limit": is this a true detection limit from the sensor/retrieval or do you mean rather than you applied a lower DU threshold (5 DU) for the NRT TROPOMI data compared to the SP ILM SO₂LH retrieval data (20 DU)? Not clear to me if this is a direct 'detection limit' or more a 'chosen threshold' based on various limitations (not necessarily a detection limit). For the 5 DU threshold you mention this is applied to avoid assimilating SO₂ from outgassing volcanoes which are covered by SO₂ emissions in the CAMS model. Also see related question below.

L416: "FP_ILM SO2LH retrieval (v3.1) does not provide reliable information for TCSO2 < 20 DU and therefore only picks up those parts of the plume that are associated with the highest SO2 load" The work 'information' is ambiguous. Does it mean that both the retrieved column load values and the layer height values are not reliable under 20 DU, or is it only the retrieved layer height data which is not reliable under 20 DU? Maybe to add in section 2.2.

L425/ Figure 9: Would be useful if the figure caption could explain why the NRT TROPOMI data differ to the SO2LH TROPOMI data (i.e., DU levels used/displayed).

L430: It is not directly explained why the SO2 burden is so much larger (2-3 Tg) for the LHexp compared with BLe xp. Is it because of higher TCSO2 values as well as overestimating the plume area? 2-3 Tg is quite a lot higher than the total burden values from the satellite data.

L575-L590: It would be good to compare these skill time scales to what was found by de Leeuw et al for the NAME model (skill for 12-17 days for the low-density (<1 DU) parts of the SO2 cloud and 2-4 days for the denser parts (>20 DU) of the SO2 cloud).

Technical comments

Figure text and labels need to be increased as on a print-out version some figures (especially figures 4, 12, 13, 16,17) are very hard or near to impossible to read.

Figure 3: Suggest changing the colour scale as there are very few values >100 DU so hard to distinguish the dots. Is this showing values only >20 DU as you mention the retrieval is accurate only for larger DU values.

The de Leeuw reference should be updated to the final revised version for 2021.

The reference Prata et al. 2019 is used in the main text but is not in the reference list.

Figure 18 could be removed as there are many figures and the difference to Fig 17 is not very big so describing by words in text should be sufficient.