Replies to reviewer 1

Thanks a lot for taking the time to read the paper and giving us valuable comments. We have changed the manuscript according to the suggestions and have listed our replies and changes in blue below.

The authors present the assimilation of SO2 retrievals from Tropomi satellite observations in the global forecasting system used in CAMS for volcanic forecasting. As for other major centres, assimilating vertically-integrated information on SO2 from space-borne sensors is a challenge which needs continuous improvement, as observational product and data assimilation settings can be refined or improved year after year. This paper is of interest to the community. I suggest it is accepted after modifications are made.

Scope and title

The title is a bit misleading as the study presented in this manuscript is presenting assimilation experiments carried out in a different system than the near real time (NRT) CAMS system used for volcanic forecasting. Moreover, the present study mainly compares results obtained assimilating the new product proposed by the DLR including information on the SO2 plume vertical extension, with several settings, to those obtained in thecurrent operational setting with NRT Tropomi data isseminated by ESA. In addition, the study described in this manuscript only focusses on a particular eruptive event, the Raikoke 2019 eruption, which injects S02 plumes at very high altitudes. No other event is assessed in this study. Eruptive events release SO2 plume at a large range of altitudes, depending on the volcano and the given episode. The present paper does not provide any guidance for other eruptive events. I suggest to change the title so as to reflect the content of the paper more closely, such as
"Evaluation of the assimilation of the S5P-Tropomi SO2 layer height product in the CAMS global system in the case of the Raikoke 2019 eruption".

We have changed the title to:

'Evaluation of the assimilation of S5P/Tropomi SO2 layer height data in the CAMS global system for the Raikoke 2019 volcanic eruption.'

Assimilation settings for the observations

Section 3.2.1 (235)
The authors describe the baseline configuration and say "SO2 observations are currently only assimilated ... when the observed SO2 concentrations are considerably larger than the atmospheric background values". I suggest the authors clearly state that criterion, instead of vaguely referring to "considerably larger".

We already mentioned in section 2.1 :' Furthermore, only TROPOMI SO2 pixels with values greater than 5 DU are assimilated in the operational CAMS system to avoid assimilating SO2 from outgassing volcanoes which are covered by SO2 emissions in the CAMS model' . For GOME-2 we assimilate all the pixels flagged as volcanic, which is also stated in section 2.1. With the statement in line 235 we only wanted to illustrate that we can not use an NMC style method because the resulting background errors would peak at the surface where anthropogenic emissions lead to the largest SO2 values. They would not give us background error statistics which would be useful for volcanic eruptions as that information is not in the model's background field. Reading the sentence again, the part 'considerably larger…' is not really needed and we have removed it, so that the sentence now simply reads:
'SO2 observations are currently only assimilated in the CAMS system in the event of volcanic eruptions.'

I may have missed the description of the observation pre-processing in the paper. Can the authors state clearly how the mismatch between the observation resolution and the model resolution? Are data thinned? Is there a super-obbing step? What are the parameters of the pre-processing?

The TROPOMI data are super-obbed to the model resolution. We already mention this in Section 2.1: 'The TROPOMI SO2 data are averaged to the model resolution (TL511, about 40km) before being used in the CAMS system. '
The GOME-2 data are used at the satellite resolution which is similar to the model resolution. We have added in Section 2.3:
'The GOME-2 data are used at the satellite resolution which is similar to the resolution of the CAMS model used in this paper.'

As the number of observations varies between NRT and LH SO2 observations, a clear indication of the difference in the number of assimilated data should be clearly given.

We already show in Figure 8 a timeseries of the number of observations and have already this text in the paper: 'Figure 8 shows a timeseries of the number of observations that are actively assimilated in both experiments, i.e. the number of 1ºx1º grid points with active observations, and illustrates that there are more active data in BLexp where NRT TROPOMI SO2 data with values greater than 5 DU are assimilated (i.e. as done in the operational CAMS system) than in LHexp where only data with LH TCSO2 greater than 20 DU are assimilated.'

No word is said on the observation errors, which are also important players in the game. The reader would benefit from a clear description on how the observation errors are handled.

We use the observation errors given by the data providers we have added a sentence in Section
' For the TROPOMI data (and also the other SO2 products used in this paper) observation errors as given by the data providers are used.'

NRT Tropomi SO2 observations are provided with averaging kernels. Are these averaging kernels used in the baseline configuration? Are SO2-LH observations provided with averaging kernels? If present, are the latter used in the assimilation? I suggest the authors clearly state all these "details".

The NRT Tropomi SO2 observations are indeed provided with averaging kernels. However, for the volcanic SO2 product the averaging kernels are simply 1 km box profiles that are used in the AMF calculation to represent typical volcanic SO2 profiles and do not provide any real information about the current eruption. It therefore does not make sense to use these in the CAMS assimilation system. There are 3 different averaging kernels provide for each SO2 column retrieval and the user can choose the product that best suits the situation. See TROPOMI ATBD for more information: https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-ATBD-SO2-TROPOMI.

We have added more information at the end of Section 2.1:

The DOAS vertical column SO2 retrieval requires knowledge of a prior SO2 profile to convert the slant columns into vertical columns. Because this profile shape is generally not known at the time of the observation and it is also not know whether the observed SO2 is of volcanic origin or from pollution (or both) the TROPOMI algorithm calculates four vertical columns for different hypothetical SO2 profiles. One vertical column is provided for anthropogenic SO2 with the prior SO2 profile taken from the TM5 CTM and three for volcanic scenarios assuming the SO2 is either located in the boundary layer, in the mid-troposphere (around 7 km) or in the stratosphere (around 15 km). These volcanic prior profiles are box profiles of 1 km thickness, located at the

corresponding altitudes. The NRT CAMS system uses the mid-troposphere product. TROPOMI SO2 data are provided with averaging kernels based on the prior hypothetical SO2 profiles (i.e. the 1 km box profiles centred around the assumed SO2 altitude for the volcanic columns). However, as these do not provide any real information about the altitude of the volcanic plume they are not used in the CAMS system. More information about the NRT TROPOMI SO2 retrieval can be found in the TROPOMI ATBD. For the TROPOMI data (and also the other SO2 products used in this paper) observation errors as given by the data providers are used.

Minor comments

line 397: data are gridded for comparison. What is the time step for this gridding: daily or hourly?

The calculation of the analysis or first-guess fields is done at the time and location of the observations in the observation operator of the model. Later, all data (obs or analysis/forecast) in a 12-hour analysis window are interpolated onto a 1x1 degree grid. We have added in section 4.1 (where we first mention the gridding):

'All the satellite data available during a 12-hour assimilation window were gridded onto a 1ºx1º degree grid….'

Figures showing timeseries are numerous and sometimes hardly legible (eg. 12, 13).

We have improved several of the figures, including Fig 12 and 13.

Figures showing maps are sometimes a bit small (eg. 5, 9)

We think the quality of Figures 5 and 9 is good enough for publication. The main point of the figures is to give an overview of the evolution of the SO2 plume and they are big enough for that.

Do the authors think showing evaluation for D+5 forecasts is relevant for such a study which shows the high sensitivity to the assimilation settings?

As the CAMS forecast system provides 5-day forecasts we think it is relevant to show them.