Open Access

Earth System
Science
Data

Discussions

# *Interactive comment on* "A multi-sensor satellite-based archive of the largest SO$_2$ volcanic eruptions since 2006" *by* Pierre-Yves Tournigand et al.

**Anonymous Referee #1**

Received and published: 27 August 2020

The submitted manuscript describes a dataset that brings together information about SO2 emissions from 11 of the largest eruptions which have occurred since 2006. This dataset collates the results from multiple satellite instruments making it a very valuable resource for future studies of these eruptions and for the validation of new tools for monitoring plumes of SO2 from volcanoes. While the retrieval schemes are not themselves new, this is the first effort to collate this type of data in one location and one simple format. The data was easily accessible and simple to download. I read and plotted the data for a number of eruptions from the dataset in IDL using an existing NetCDF reader and using the manuscript and supplementary information for guidance. For the most part I found this to be straightforward. Having said that, further work is

required to improve the clarity of the paper so that the data can be easily read and interpreted by the users. In particular, further thorough proof reading is encouraged to improve the readability and there should be more discussion of the uncertainties and strengths/limitations of each tool so that someone using the data interprets it correctly. There are also some inconsistencies in the dataset that should be improved. See below for the specific and technical changes suggested for the manuscript and the dataset.

Specific Comments

MANUSCRIPT

Line 17. I feel it would be beneficial to elaborate here on there being no other archive which compiles the results from multiple satellite instruments and eruptions to really emphasise the strength of this dataset.

Line 20. 'We've archived and collocated . . . the vertical backscatter from CALIOP. . .' – I think rather than 'vertical backscatter' this sentence should indicate that you have included the CALIOP height and aerosol type.

Line 25. Here you state that 'the cross-comparison of the datasets shows the high consistency of the parameters estimated with different sensors and algorithms'. This feels like quite a strong statement. In section 5 you compare the heights obtained with RO, CALIOP and IASI. You note that for a number of eruptions there is a good agreement between RO and CALIOP but that this is not the case for Calbuco. Table 4 shows that a number of the average differences between IASI/CALIOP and IASI/RO are greater than 3 km. Additionally, you have done no quantitative comparison of the partial column densities from AIRS, IASI and GOME-2. Some rewording would improve this statement.

Line 91-102. This paragraph is a little confusing. You cite three papers/datasets: Ge et al. (2016); Carn et al. (2017); Carn et al. (2019) and it is a little difficult to tell if and

how these papers/datasets are connected. Also, you suggest that Carn et al. (2017) included 'passive degassing' and 'main eruptive events' (line 94) but to the best of my knowledge this paper generates long term averaged fluxes that exclude large eruptive events. I would advise some rewording of this paragraph and perhaps some expansion on what is included in the Ge et al. (2016) and Carn et al. (2017) papers which might to help the reader better understand their content.

Section 2. In this section it would be useful to have some more information about the performance of each technique. For example, conditions in which the technique performs well or badly. And information such as the detection limits and uncertainties. This has been done for AIRS (lines 135-138) and something similar for each instrument/technique would help the reader appreciate the strengths and limitations of each tool. It would also help a user to correctly interpret the archived data- especially if they are comparing the results from different instruments. Section 4.3 does point the reader to some of the relevant literature but it would be nice to have this in section 2 and with more detail.

Line 144. You mention the IASI retrieval technique is based on a BTD with the v3 absorption band – brightness temperature difference with what?

Line 144-149. Initially it is implied that the IASI VCD retrieval is run using fixed heights. But in the archived data there is only a single value for the IASI VCD. Could you clarify if this is obtained by interpolating the results with the height from the second retrieval?

Line 148. It would be useful if there was a line here explaining how the IASI height retrieval worked.

Section 2.4. I think in this section it is important to highlight that with CALIOP you are not measuring SO2 but ash, sulphate, smoke and/or dust. It would be good to acknowledge here some of the limitations of assuming SO2/ash are collocated.

Section 3. This section details all the variables contained in the files. I think it would be

really beneficial to a user to have these listed in a table (either in this paper or in the supplementary information). I found I referred to the supplementary information (print out of all the variables) a lot while trying to load and plot the data. A table summarising the variable names, meaning, dimensions, type and units would be even more useful as a quick reference guide.

Line 201. This is the first instance that 'granule' has been used. Please can this be clearly defined. The use of the word granule made it challenging to interpret the data structures described in sections 3.1-3.3 independently of reading in and looking at the data.

Section 3.4, Section 3.5. In these sections you mention collocation with between CALIOP, RO and the other instruments. I think it would be useful to know what conditions you use for the collocation here rather than in sections 4.1 and 4.2. This would help the reader immediately understand what is meant by collocation.

Lines 340-344. Here you discuss the average differences between the cloud heights for different eruptions. Are there any reasons why the average difference is greater at Calbuco than for Eyja, Kasatochi and Grimsvotn? Also, is there a reason why the differences are greater between IASI and RO/CALIOP? I think it would be useful for a user of the dataset to understand why differences might arise between the different datasets (e.g. the time difference between the overpasses and the method used to obtain the height information). Additionally, have you considered a quantitative comparison of the VCDs retrieved with AIRS, IASI and GOME-2? What differences would you expect to see between these?

Table 1. You could add the eruption VEI and the eruption end date or duration to this table. Additionally, it could be helpful to add the geographic region considered for each eruption and the start/end date for the data in the archive – both of these would be valuable to the data user.

Table 2. In addition to the information given you could also mention the spectral

range/resolution of the instruments.

Figure 2. It would be interesting to see the cloud top heights obtained with CALIOP and RO in this plot rather than just the tracks/points.

SUPPLEMENT AND DATA

NULL values – Throughout much of the dataset the null values are reported as -9999. However, for the RO profiles they are recorded as NaN. For the RO cloud top heights it goes back to -9999. For the CALIOP heights there are no -9999 or NaN instead there are 0's- are these null values too? This should probably be consistent and whichever is chosen should be clearly noted somewhere.

Different number of variables - The files do not contain consistent numbers of variables. For example, in the file 'Calbuco_2015_05_24.nc' there is data available for IASI and RO but not GOME-2, AIRS or CALIOP. Presumably this is related to the availability of the data. It would be good to clarify this in the manuscript (perhaps at the start of section 3). Even better would be to summarise how many days or which days are covered by each instrument for each eruption – this could be an addition to table 3 and would be slightly easier to interpret than the number of granules.

Dimensions - a list of dimensions is given on page 1 and page 8 of the supplement. It would be very helpful if these were expanded on. In particular the definitions for 'CALIOP_char' and 'CALIOP_char2' are not very informative.

P1, P2 supplementary info – there is a slight discrepancy between the long names between IASI and GOME-2. For GOME-2 the long name states that the data is a composite of GOME-A and B – is it also the case for IASI that the data is a composite of IASI-A and B?

IASI_SO2 – I suggest expanding the long name of this variable to make it clear that this is a vertical column density and to explain what interpolated is referring to

P3, P4. It is not clear what the dimensions should be here. CALIOP_CHAR,

C5

CALIOP_char2 and CALIOP_type should be more clearly defined in the supplementary data. CALIOP_type (the dimension) is not defined in the dimensions list.

CALIOP_type – This variable was very challenging for me to read in (in both IDL and python). The supplementary information (page 4) suggested that these were doubles but they had to be read in as strings. I think the choice of saving these as a string is so that multiple flags can be indicated. Initially on reading in this variable I obtained an array with 3 dimensions. These then had to be converted to strings and joined together to extract the CALIOP type (a similar thing had to be done for CALIOP_filename- also not immediately obvious how to read in IDL). Following that the newly joined strings had to be searched to determine which aerosols were present. Could there be a better way of saving this variable? Perhaps simply an integer array for each variable type with 1 indicating the presence of this aerosol and 0 indicating its absence. Alternatively, more information on how to read in and interpret these results would be very useful.

P5-P7. For the RO variables expanding the long names for 'air_temperature', 'air_pressure', 'refractivity', 'specific_humidity' would provide more information- these could for example mention that these are profiles.

RO – cloud top heights. The units do not seem to be consistent for these (in the daily files). For colocations with AIRS and IASI the heights appear to be in meters (which are the standard units and consistent with heights reported by CALOP and IASI). Whereas for GOME-2 they seem to be in km.

P4-7. The dimensions for the RO profiles are listed as RO_AIRS_lat by RO_AIRS_PROFILE (or IASI/GOME). Could these be defined more clearly in the dimensions list.

Dates covered by each eruption. Some of the daily files start before the start date of the eruption. For example, for Nabro (eruption starting on the 13th June 2011) the first file in the dataset is 31st May 2011. In the first few files it seems to include the outputs for other eruptions. For example, the file Nabro_2011_05_31 includes SO2

C6

measurements from the Grímsvötn eruption, while the file Nabro_2011_06_05 includes measurements from both Grímsvötn and Puyehue. Including this twice in the dataset is a little unnecessary and means the user has to download more data than is needed for this eruption. It is possible to see plumes from different eruptions in many of the datafiles.

Technical Comments/Suggestions

MANUSCRIPT

Throughout – Some of the volcano names have accents (e.g. Grímsvötn, Eyjafjalla-jökull, Puyehue-Cordón Caulle)

Line 16. 'Forecast' should be forecasting or forecasts

Line 17. 'Single events' would be more precise as 'single eruptive events'

Line 17. '. . . but not any archive is available' need rewording. Perhaps: '. . . no such archive is available'.

Line 18. 'from three different instruments' would be clearer as 'from three different satellite instruments'

Line 19. 'the atmospheric parameters vertical profiles from. . .' This line is a little confusing. Perhaps rephrasing as something like: 'vertical atmospheric profiles obtained from. . .'

Line 21. 'We additionally' would read better as 'Additionally we'

Line 22. 'The dataset consists of 223 days monitored with SO2 clouds' This line does not read very well – consider rephrasing it.

Line 38-39. What is meant by 'consequent cloud'? – are you referring to the volcanic cloud or ice/water clouds (e.g. indirect climate effects)

Line 40. 'SO2 injections in the stratosphere' may read better as 'SO2 injections into

the stratosphere'

Line 42. 'hence transported' may read better as 'hence be transported'

Line 46. 'has occurred per year since 1994 worldwide' might read better as 'have occurred worldwide each year since 1994'

Line 47-48. '. . . the energy of the eruption, amount, type and size of the ejected material' would read better as '. . . the energy of the eruption, and the amount, type and size of the ejected material'

Line 49-51. To improve sentence clarity move the Newhall and Self reference to the start of the sentence: 'The VEI was introduced in 1982 by Newhall and Self (1982) . . .'

Line 50. I think it is Richter scale rather than Richter's scale.

Line 50. I think it should be earthquake rather than earthquakes'

Line 54. 'VEI index' can just be VEI

Line 60. Putting 'e.g. VEI 4 events' within brackets would help the readability of the sentence

Line 71. 'and' should be used instead of 'or'

Line 72. 'although' would make more sense than 'even though'

Line 74. 'focusing on single or a few eruptions' would read better as 'focusing on a single or a few eruptions'

Line 77. Stating that 'all' platforms and algorithms were studied in this volume seems quite strong. Perhaps: 'a large number' would be better

Lines 77 and 81. Starting the sentence with 'Sarychev Peak 2009' and 'Grimsvotn 2011' does not read very well. It might sound better as 'The Sarychev Peak eruption in 2009. . .' etc.

Line 91. '. . . and updated in the course of the years.' This line does not read very well – consider rephrasing.

Line 110. It should read '... and humidity from GNSS RO profiles'

Line 111-112. This sentence would benefit from being rewritten to improve the clarity. Maybe something like: 'This information is provided for eruptions, after 2006, classified by the GVP as VEI 4 or larger and with an SO2 mass loading of greater than 0.05 Tg. At the time of archive preparation, no eruptions after 2016 had yet been classified as VEI 4 or greater.'

Line 113. Rather than include '(table 1)' in this sentence, perhaps add a sentence at the end of the paragraph saying 'Further information on these eruptions can be found in table 1.'

Line 117-118. 'there is no current unique database'. This does not read very well – I would suggest rewriting the sentence

Line 119-121. It should read 'accurate knowledge of volcanic SO2 cloud concentration and altitude as well as their spatial and temporal evolution. . . of an eruption's climatic impact'

Line 122. 'retrievals' should be 'retrieval'

Section 2 – title. Maybe this should be titled 'Instrument and Retrieval Description'

Line 126. 'due to their own limitations.' It is not clear what is meant by this.

Line 130. It should read 'an ascending orbit'

Line 133. It would be good for a reference to be included for this sentence so the reader is immediately aware of which paper describes this technique.

Line 135. Again it would be good to have a reference for this statement.

Line 142. It should read 'an ascending orbit'

Line 144. V3 has not been defined. SO2 is also not formatted correctly.

Line 151. Slight inconsistency - On board has a space here but elsewhere it is written onboard.

Line 153. Slight inconsistency - Here the pixel size is listed as 40x80 km. For AIRS it was written as 13.5 x 13.5 km (with spaces).

Line 165. 1,67 should be 1.67

Line 175-176. It may read better as 'In this archive we use the RO bending . . .'

Line 192. It should read 'the number of days' rather than 'amount of days'

Line 232-233: 'Four of those types are of interest for this archive: type 2, 6, 9 and 10 respectively corresponding to dust, elevated smoke, volcanic ash and sulfate/other.' Include a colon between archive and type.

Line 234-235. There should be a space between 20.2/30.1 and km

Line 240. I think a colon would be better than a comma between provided and latitude

Line 264. Should read 'Where $\alpha$ is the bending angle anomaly. . .'

Line 277. Should read 'consists of' rather than 'consists in'

Line 285. I think this should be 'Thus we' rather than 'We thus'

Line 308. Having a list (in brackets) of parameters that affect the uncertainty, followed by a line about altitude affecting the uncertainty does not read so well. Maybe combine all the factors that affect the uncertainty into one line.

Line 315. Should read 'volcanic cloud' detection rather than 'volcanic clouds'

Line 317. It should read 'in charge of processing them'

Line 317. Is it not 10 VEI 4 and 1 VEI 5 eruptions? The Puyehue eruption in 2011 is listed on the GVP as VEI 5 (https://volcano.si.edu/volcano.cfm?vn=357150 ; under

eruptive history). Also here you state the period you a looking at is 2008 to 2016 when previously you've said you were looking at eruptions from 2006-2016.

Line 318. 'With a total of' rather than 'for a total of'

Line 320-321. 'Several parameters are measured using different instruments, such as SO2 VCD and cloud top altitude, to allow cross correlation between the different retrieval algorithms.' Do you mean to say 'Several parameters are included within the dataset. . . to allow cross correlation between the different algorithms' ?

Line 335. It should read 'compared the date, time . . .'

Line 336. 'We have additionally' would read better as 'Additionally we have'

Line 360. Not just detection but also the retrieval of VCDs

Line 365. 'Up to date' would read better as 'At present'

Line 370. 'and test new algorithms contributing to improving the accuracy on the estimation of fundamental volcanic clouds parameters'. This may read better as 'and test new algorithms on, thereby contributing to improving the accuracy on the estimation of fundamental volcanic clouds parameters'

Line 373. 'allowing to reconstruct. . .' may sound better as 'allowing the reconstruction of. . .'

Table 2. Maybe differentiate between AIRS and IASI spatial resolutions (13.5 by 13.5 km vs. 12 km diameter circular pixels)

Figure 2. In the caption 'upright' should be 'top right'. Also, this caption reads a little strangely. I would suggest: 'Example of data use and data collocation. (a) Kasatochi cloud on 9th August 2008; (b) Sarychev peak cloud . . .' At present there is no (b). Additionally, no full stop is required in line 610.

SUPPLEMENT AND DATA

P1. VC – is undefined in the supplementary material.

P1. DATE_IASI – The use of the word 'because' in this description doesn't make sense.

P2, P9. In the GOME_lon variable- dimensions include GOME_late rather than GOME_lat