General remarks:

We thank the reviewers again for making very useful suggestions to further improve the paper. Our point-by-point responses to the reviewers' comments and corresponding changes are detailed below in blue text, and the changes are shown in the version of the manuscript with tracked changes (referring to the Manuscript from 16 Feb. 2022 and the last request for major revisions from 27 May 2022).

We thank the editor A. Schmidt and the reviewer for pointing to some misunderstandings concerning the different treatment of background stratospheric aerosol and minor eruptions in EMAC and WACCM. We have corrected the text in the introduction, section 6.3 and conclusions accordingly.

GloSSAC is included now in most frames of Figures 9 and 10 using provided extinction. An additional figure using GloSSAC as counterpart to our Figure 8 is provided in Appendix C1.

We have added also remarks on possible underestimates in radiative forcing due our calculation method in EMAC. The discussion during the review process has led to significant model improvements:

Due to a new much more powerful super computer we were able to redo our simulation with an improved code which considers the contribution of aerosol in the extratropical lowermost stratosphere to radiative forcing as indicated in the reply to the previous version. Now the results are much more consistent with other studies. Figure 11 contains now the improved method for the calculation of forcing. Section 6.3, section 3, abstract, conclusions and appendix C3 are adapted accordingly.

Figure 11 had to be redesigned. We show now mostly the forcing at the top of the atmosphere but kept for the previous versions the published values at the tropopause (see also corresponding TOA values in the supplement):

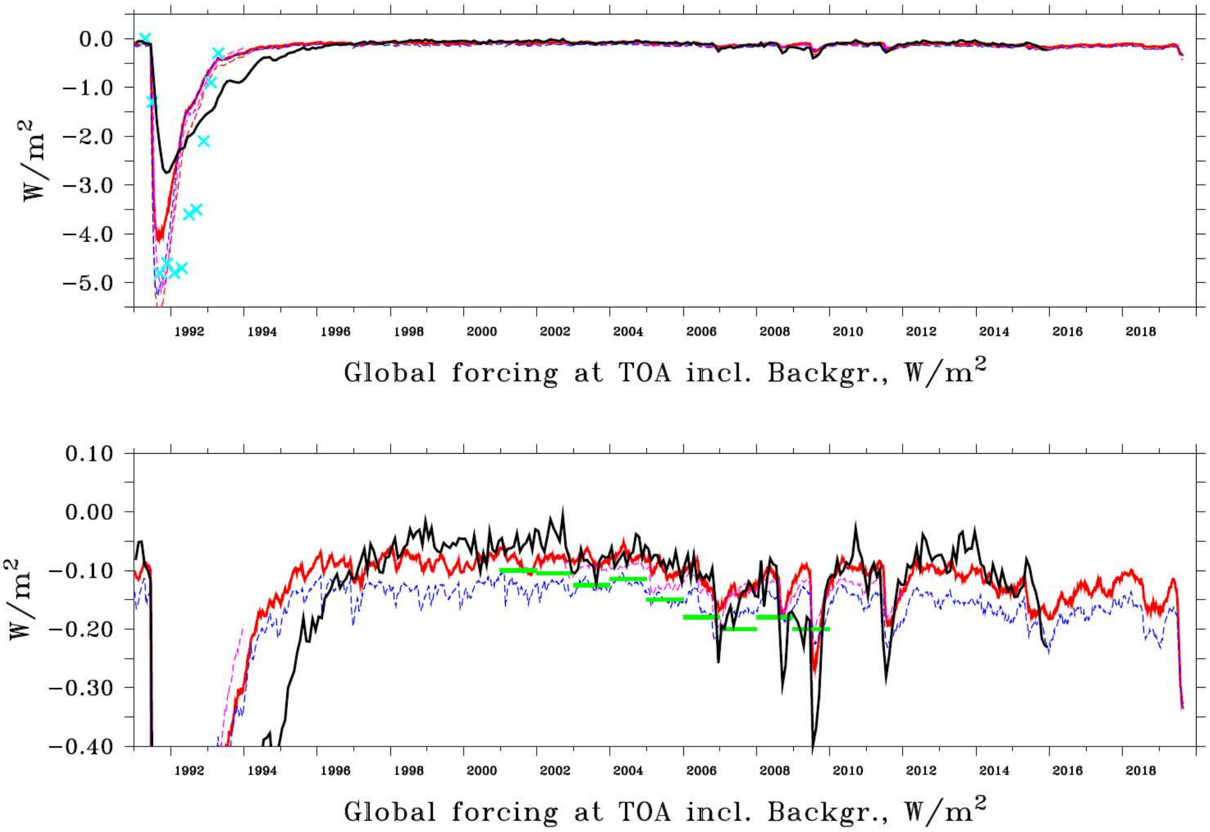

Figure 11. EMAC instantaneous radiative forcing by stratospheric aerosol (red, pink and blue lines, 5-day averages). Solar forcing at the top of the atmosphere (TOA) (dashed red line, upper

panel) is compared to solar forcing at the TOA from satellite observations of the Earth Radiation Budget Experiment (ERBE), 72-day means, light blue crosses (Wong et al., 2006; Toohey et al., 2011). The full red line displays the total (solar+IR) forcing at TOA including contributions from aerosol down to the calculated tropopause. The blue dashed line shows the total forcing using the old scheme at the tropical tropopause (Brühl et al., 2018), the dashed pink line the same with less volcanoes of Brühl et al. (2015). Green bars show annual averages derived from observations by Solomon et al. (2011). The black line shows results from Schmidt et al. (2018) with volcanic radiative forcing at TOA including a background aerosol forcing of -0.05 Wm^{-2}. The lower panel is a zoom of the upper panel.

Abstract: "These eruptions cause a total global radiative forcing of the order of -0.1 Wm^{-2} at the top of the atmosphere (TOA) compared to a background stratospheric aerosol forcing of about -0.04 Wm^{-2}.
Medium strength eruptions injecting about 400 kt SO_2 into the stratosphere or accumulation of consecutive smaller eruptions can lead to a total forcing of about -0.3 Wm^{-2}. We show that it is critical to include the contribution of the extratropical lowermost stratospheric aerosol in the forcing calculations."

Conclusions: End rearranged. Last paragraph replaced by: "Note that for the calculation of the forcing by these medium size eruptions it is essential to consider the radiative effect of volcanic aerosol down to the tropopause (Ridley et al., 2014). Including only the aerosol above the 100 hPa level as done in Brühl et al. (2018, 2015) can lead to significant underestimates which were partially hidden in these studies by showing the forcing at the tropopause which is about 0.08 Wm^{-2} stronger than that at the TOA in EMAC."

Appendix C3: "Five-day-average radiative forcing at the TOA is shown in Figure C4. The black curves correspond to the red curves in Figure 11. Using the assumptions of Mills et al. (2016) on the vertical distribution of the SO_2 injection leads to less forcing in case of Nabro and after about 5 weeks after the Sarychev eruption (Figure C4, red curves). For the latter the peak is stronger since more mass is injected than in the other approaches. It is dominated by contributions from aerosol in the mid and high latitude lowermost stratosphere. Using Table 2 leads to approximate agreement if the provided altitude is about the top of the column into which is injected, using as a bottom about 1.2 km above the latitude dependent minimum altitude (green curves). Using the full range leads to an underestimate compared to the simulation using the MIPAS observations directly as 3-D SO_2 perturbation (blue curves). This is due to more efficient removal processes in the upper troposphere and the lowermost stratosphere than in the layers above. The lower panel for 2009 contains also a sensitivity study similar to the green curve but with the eruption of Mando Hararo neglected as done by Mills et al. (2016) (purple), and one where the top of the injection column for Sarychev was one layer lower than in the case with the blue curve (light blue). These examples show that the altitude of the injection has a large impact on the radiative forcing, but also that not only medium size eruptions matter. For stratospheric AOD shown in the upper panels of Fig. C4 the difference between the approaches is less than for the forcing in the first weeks after the eruption. Later the SAOD decreases faster with the point source approach, especially the one of Mills et al. (2016), than with our method.
Note that considering only aerosol above the fixed pressure level of 100 hPa for the forcing calculations as we did in earlier studies causes misleading results here."

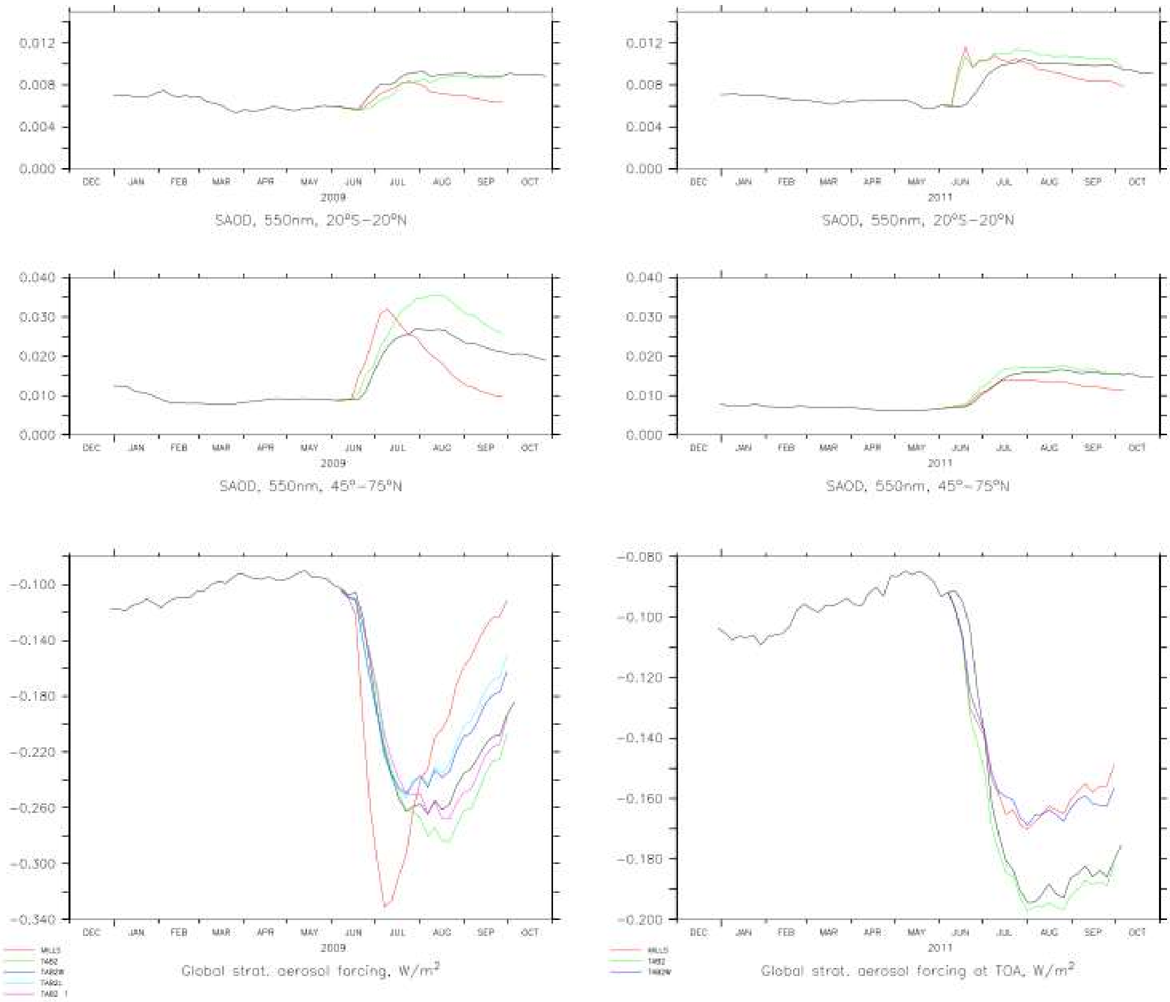

Figure C4. Stratospheric AOD (upper) and global radiative forcing at TOA (lower), Sarychev (left) and Nabro (right), black line as red line in Figure 9 or Figure 11, blue, light blue and green lines "point sources" based on Table 2 with different thickness and vertical position of the column into which is injected (see text), red lines with assumptions of Mills et al. (2016). Purple and light blue curves see text.

We checked the whole manuscript and corrected or removed misleading or distracting words and sentences, including minor corrections in the caption of Figure 7.

Report #1

Anonymous Referee #1

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

All line numbers refer to tracked changes version of manuscript.

L52: "Radiative forcing is…" I don't believe it is true that this approach is "valid only for purely scattering aerosol". The -25 factor comes from Hansen et al. (2005), from model simulations including prescribed stratospheric aerosol for Pinatubo, and is based on the TOA net (SW+LW) flux anomalies, thus incorporating both scattering and absorption effects.

-> Response: Text modified: "…valid only for scattering sulfate aerosol (see e.g. Sellitto et al. (2022))." There are several other references on that, nevertheless, the factor is still good for plausibility studies.

L56: The description of GloSSAC is somewhat improved compared to prior versions. But still, some evidence (via citation) should be included to support the claim that GloSSAC includes "large uncertainties due to data gaps"? I think here you are really referring to the issues with "saturation" after Pinatubo? If so, it would be helpful to be specific about it and make reference to statements in the relevant literature, e.g., from the GloSSAC papers.

-> Response: Rephrased to a more neutral formulation without the remark on data gaps: "The model simulations in this study are compared to GloSSAC V2 (Thomason et al., 2018; Kovilakam et al., 2020), a time dependent multi-satellite zonal average aerosol climatology which provides extinction data (Figure C1 in Appendix C1)".
The gap-filling in the Pinatubo-period is addressed in detail in section 6.2 based on Thomason et al 2018. (see also below).

L57: It would still make sense to include a mention of SAD here since you use it in your analysis.

-> Response: SAD is taken directly from the Level2-SAGE data, in GloSSAC V2 it is not provided. Remarks on that in the MS were removed.

L202: How are SSTs treated, interactive or prescribed? This is important for the calculation of radiative forcing and the comparison with the simulations of Schmidt et al. who used prescribed SSTs.

-> Response: Section 3 is expanded. We mention now ECHAM and inserted:
"The temperature and the dynamics above the boundary layer are nudged to the meteorological ERA-Interim reanalysis data of the European Centre for Medium-Range Weather Forecasts (ECMWF) up to about 100 hPa, while the sea surface temperatures (SST) and sea ice are prescribed using ECMWF data (more details in Jöckel et al., 2006)".

In section 3 (line 232) we adapted the text for the new calculation method of radiative forcing: "… instantaneous forcing is calculated online from the difference of fluxes for the cases with stratospheric aerosol above the calculated tropopause (in earlier studies only above 100 hPa) …"

L263: The correction factor is still unclear to me. I appreciate more detail being added to the appendix, but the description in the main text needs to be understandable enough for the reader to make sense of what it is. It would help in the first sentence to clearly state what quantity the correction factor is applied to and what the target is. The 2nd sentence here says "we iterated calculated extinctions to agree with OSIRIS", so this sounds like the correction factor was applied to the extinctions from other instruments (GOMOS?) to produce best agreement with OSIRIS extinction. But later it is mentioned for one case that a correction factor of 0.8 is applied to OSIRIS? Is it not rather that the correction factor is applied to the non-SO2 measurements (OSIRIS, SAGE and GOMOS) to produce best agreement with MIPAS SO2? Or is the correction factor also applied to MIPAS in some cases? Is the factor 1 for all cases when the target dataset (MIPAS or OSIRIS?) is not available?

-> Response: We added the formula for the SO_2 mixing ratio perturbation to the main text in section 4:

"The SO_2 mixing ratio perturbation ΔVMR is derived from the extinction perturbation $\Delta\beta_{ext}$ (750nm) as in Equation 1 using a constant ratio between model calculated sulfate concentration and its share on extinction in the lower stratosphere of low latitudes:

$$\Delta VMR = 1.2 \times 10^{12}\, \Delta\beta_{ext}/\rho \;\; f \qquad (1)$$

ρ is the altitude dependent air density in molecules/cm^3 (Examples and more details see Appendix B). We assume that the spatial patterns of the perturbation of extinction and sulfate are the same as for SO_2. A similar technique is used for OSIRIS and can be used for SAGE II data.

If data gaps cause a shift of the time period away from the maximum perturbation or a bias in the zonal average, a correction factor ($f \neq 1$) is applied to Equation 1."

Line 605 (Appendix B) is expanded for clarity: "If the time lag of data is several weeks a correction factor >1 has to be applied in Equation 1 to account for removal processes, if another event is relatively close in time, the factor has to be <1 to remove the influence of the previous event (see Table S1 in supplement, which indicates also additional uncertainties for a few cases in 2018 and 2019 due to too sparse data)."

Added in line 247: "For MIPAS sometimes corrections in the order of 30% were necessary because of gaps. Here the corrected values serve as reference for the other instruments".

More details see next point.

L266: In this paragraph we have statements that the correction factor takes values "as high as 3" and "up to 2" with no clear indication of a difference in scenarios between those two maximum values. Please clarify.

-> Response: We modified the text to clarify, that the correction factor '3' and the following sentence refer only to Calbuco:

"… Correction factors up to 2 have to be applied in some cases because of data gaps, incomplete profiles (both containing zero values) or for high latitudes (examples see Appendix B). One exception is the eruption of Calbuco, with a correction factor of 3 for removal processes, because of a shift of three months due to a big data gap. To estimate the factor in this worst case, we iterated calculated extinctions to agree with OSIRIS and also used observations and assumptions by Vernier et al. (2016) like the decay of extinction by sulfate with time over 4 months. On the other hand, the factor can be as small as 0.5 to account for sulfate remnants of eruptions occurring 2-4 weeks before the date of the eruption to be analysed or for cloud perturbations. These factors, together with the used time periods, are provided in the electronic supplement (Table S1 for OSIRIS and Table S2 for GOMOS)".

L304: What does "boxes related to the volcano" mean?

-> Response: This paragraph is rephrased and moved to the previous section because this describes the quantities provided in the table, but not how we use the data in the simulation. See point L306.

L305: "split into boxes considering the mean wind in the lower stratosphere and consistency with nadir observations" does not help understand how or why this split is done.

-> Response: Rephrased and rearranged, see below.

L306: “lower boundary…” this also does not make sense to me. Is this related to the different min altitudes for tropical and extratropical eruptions that are mentioned later in the manuscript? It wouldn’t seem to make any sense to use a lower integration limit over the whole globe for an extratropical eruption. In any case, here you are talking about a single case, so why not just mention what it is and how the total SO2 was calculated?

-> Response: Moved to previous section and rewritten, to make it clear that it refers to latitude belts. This is corrected also in the caption of Table 2.
In section 4: “The amount of sulfur emitted by each eruption is calculated by integration over the three-dimensional SO_2 perturbation plumes, excluding tropospheric emissions below 12 km at high latitudes, 13 km at mid-latitudes, and 14 km at low latitudes. The latter is selected to include possible convective transport from the upper troposphere into the stratosphere in the tropics. The limits in mid and high latitudes above the mean tropopause were selected to exclude cloud perturbations by frontal systems or uncertain satellite data. This can lead to an underestimate of injected mass in some cases. The plumes don't cover the whole globe, they are always in a latitude range derived manually from the satellite data.”
Table 2: ”… integrated over latitude belts above 14 km … high latitudes from the 3D mixing ratio perturbations. Listed altitudes and latitudes represent the region of maximum mixing ratio perturbation, the altitudes are close to the top injection height.”

L341: Please rephrase to remove “only if”—I doubt all other possible methods have been attempted.
-> Response: Sentence rephrased to remove “only if” by "by considering". Added: “… or most other data bases in ISAMIP, e.g. Mills et al. (2016).”

L391: This statement is false: tropopause heights can be extracted from any meteorological reanalysis, and are often available as part of satellite data sets since temperature is typically an ingredient of the retrieval method.

-> Response: Thanks for the remark. We removed that statement. We argue now that fixed heights are more convenient here for comparison with existing literature.
“For practical reasons, the total stratospheric Aerosol Optical Depth (AOD) is obtained by the vertical integral of the aerosol extinction above an altitude of about 16 km in the tropics and above about 13 km for mid-latitudes and high latitudes, to allow for a direct comparison with existing literature (Santer et al., 2014; Glantz et al., 2014) and satellite data.” See also captions of Fig. 9 and 10.

Furthermore, we have improved the radiative forcing calculations by using a model calculated tropopause height, instead of using a fixed pressure level of 100 hPa (see General remarks and sec. 6.3).

L399: “Note the odd…” There is nothing odd per se about the downward trend in GloSSAC AOD in 2012, the slope of this trend is similar to other periods including 2016-2017 and 2007-2008. The OSIRIS data mentioned that doesn’t show this downward trend—do you refer to Fig 10? To my eye, both OSIRIS and GloSSAC show a mostly monotonic decrease from 2012 to 2014, reaching a minimum value which is slightly lower than any time before around 2005. So qualitatively, I don’t see anything “odd” about GloSSAC compared to OSIRIS in and after 2012. What is clear is that there is a discrepancy between GloSSAC and the model beginning in 2012, which is not apparent in the comparison between OSIRIS and the model at 750 nm.

-> Response: We have skipped the sentence to avoid confusions. C.Br. contacted the responsible scientist for that at the SSIRC-Meeting in Leeds. The problem is visible also for 750 nm if derived from 525 and 1020 nm GloSSAC data, but less. GloSSAC AOD is now in the upper 2 panels of Fig. 9 and 10, integrated over the same height ranges as for the other data from extinction.

L405: GloSSAC doesn’t try anything, it’s a dataset.

-> Rephrased: “In GloSSAC gap filling (with lidar and CLAES data) was applied for this case.”

L405: Larger than what?

-> Response: Sentence removed.

L415: What temporal resolution is the EMAC simulations shown in? This should be stated in the figure caption. Comparisons should be made at the same temporal resolution--just calculate monthly means of the EMAC data for a comparison with the Schmidt et al. simulations.

-> Response: The EMAC simulations are based on 5-daily emission data sets from MIPAS, GOMOS and OSIRIS. "5-day averages" or "monthly data" is added to the figure captions. Typical numbers of the temporal resolution effects are mentioned in the text of section 6.2 and 6.3. Text modified, figures with a comparison are provided in the supplement (Fig. S2 and S3).

L446: There are many other differences between these simulations, they use different models! This is not enough evidence to make such a confident statement.

-> Response: Text of section 6.3 modified. There was a misunderstanding concerning the treatment of background stratospheric aerosol. The model output of the values was set to the level of TOA and a background forcing taken from their paper was added to the forcing of Schmidt et al. (2018) (Eq.1, first term, which appears to be closest to our approach since we don’t consider aerosol cloud interactions) to avoid to compare apples and oranges. See also general remarks.

“The new model simulations for the forcing at TOA with the additional volcanic eruptions (red line) are closer to the estimates from satellite extinction measurements of SAGE, GOMOS and CALIOP by Solomon et al. (2011) (green bars) than in previous studies (e.g., Brühl et al. (2015), pink dashed line and Brühl et al. (2018), blue dashed line, for TOA see supplement, Fig. S1). In Fig. 11 the published forcing values at the tropical tropopause of the previous studies are shown which are systematically more negative than the values at TOA. For Sarychev it is clear that this cannot compensate for the effect of the neglect of aerosol in the extratropical lowermost stratosphere. A comparison with volcanic radiative forcing (aerosol-radiation interactions) from Schmidt et al. (2018) is shown by the black line, including an instantaneous forcing of -0.05 Wm^{-2} for stratospheric background aerosol (derived from numbers provided).

Especially for high latitude eruptions their forcing is larger than EMAC and the annual averages of Solomon et al. (2011), also because of a higher aerosol load in the lowermost stratosphere than in EMAC (see sensitivity studies for Sarychev in Appendix C3)."

L449: "Dominant factor" implies a comparison between different factors—what are you comparing to?

-> Response: Text modified: "In the period considered here, the volcanoes are the dominant factor in instantaneous global radiative forcing. Background stratospheric aerosol like sulfate from other sources, dust and organics contributes about -0.04 Wm^{-2} to the value in the order of -0.1 Wm^{-2} at TOA in volcanically quiescent periods (e.g. in 2000, 2002 or 2004). At TOA absolute values up to -0.2 Wm^{-2} (-0.14 Wm^{-2} old approach where only aerosol above 100 hPa was considered) are reached after Rabaul (2006), Kasatochi (2008), Nabro (2011) and Calbuco/Sinabung (2015) and stronger than -0.32 Wm^{-2} (-0.2 Wm^{-2} old approach) after Raikoke/Ulawun (2019) eruptions. The value for Raikoke/Ulawun is within the range discussed in Kloss et al. (2021)."

L450: concerning the RF values given here, are these absolute values of the simulated IRF, or anomalies with respect to the 2002 quiescent period? This needs to be explained more fully. The sentence seems to imply anomalies, but the values look like absolute values. (The peak after Raikoke is around 0.3 W/m^2, and the value in 2002 is around but larger than 0.1, so the difference should be less than 0.2?). Also, what do you mean by "for" when you connect these values to certain eruptions? It seems clear that as you say, the RF depends also on the history of eruptions before, so it would be better to say the IRF reaches particular values "after" these eruptions, so as not to imply these values are fully attributable to those specific eruptions.

-> Response: Thanks for the remark. The given values are absolute (or total) values. The values have been standardized to total values throughout the text for TOA and tropopause. All values of radiative forcing have been corrected by the results of the improved simulations with aerosol above the calculated tropopause considered.
Corrected formulation "for" -> "after". See above, point L449.

L459: Schmidt also included nudging of meteorology. Critical is also that Schmidt used historical SSTs and sea ice.

-> Response: Here the models are similar, the main difference is that Schmidt et al. (2018) used two simulations and we just one with diagnostic output of radiation fluxes. A remark on that is included now in section 6.3.
In EMAC as described in section 3, see point L202.

L513: Better than what?

-> Response: Wording modified. "an improved SO_2 emission database"

L533: "These cause a global negative radiative forcing of 0.12 (0.22 to 0.08) W m^-2": what does the value 0.12 Wm^-2 mean here, do you mean here the value of RF during the quiescent period, or an average over some other period? What does the range 0.22 to 0.08 mean? Later in the sentence, values are given which don't seem to match what was given in Sec 6.3 (0.2 or 0.3 Wm^-2 "for" Raikoke?). Coming here as the last sentence of the conclusions, statements should be repeating or summarizing results shown earlier in the paper, not bringing up newly derived values.

-> Response: Thanks for the remark. The values have been standardized to total values throughout the text for TOA. This part is now consistent with section 6.3.
“These cause a global negative radiative forcing of the order of more than 0.1 Wm^{-2} at the TOA, including a background aerosol forcing of about 0.04 Wm^{-2}. For example, in the case after the eruptions of Soufriere Hills/Rabaul (2006), Nabro (2011) and the combination of the Sinabung, Wolf and Calbuco eruptions (2015) a negative radiative forcing of down to 0.2 Wm^{-2} (0.14 Wm^{-2} old approach), and 0.32 Wm^{-2} (0.2 Wm^{-2} old approach) was reached after Raikoke/Ulawun (2019) at TOA.”

Fig B1 and B2: please give units for the quantities shown on the plots.

-> Response: Added "ppb" in the captions and "Altitude, km" on y-axes in the figures.