Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2022-514-RC1, 2022

Comment on acp-2022-514

Thomas Aubry (Referee)

Referee comment on "Interactive stratospheric aerosol models' response to different amounts and altitudes of SO_2 injection during the 1991 Pinatubo eruption" by Ilaria Quaglia et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-514-RC1, 2022

General comment

This manuscript presents some of the first results stemming out of the Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP). There are large discrepancies among climate model with interactive stratospheric aerosol capabilities, and they are not successful in reproducing in details the response to the Pinatubo 1991 eruption, the only well-observed large-magnitude eruption to date. The ISA-MIP HErSEA experiment analyzed in this paper provides new insights on how uncertainties regarding both the Pinatubo emission characteristics and model uncertainties affect our capability to understand the evolution of stratospheric aerosol properties following large magnitude eruptions. A number of important results are presented, in particular that none of the participating models can reproduce the aerosol lifetime in the stratosphere, and that most of them exhibit a too strong transport from the tropical to the extra-tropical stratosphere.

Overall this is an excellent study, perfectly adequate for ACP, and I recommend it for publication after addressing moderate and minor comments listed below. In particular I strongly suggest that sensitivity experiments including the Hudson 1991 eruption are performed, for at least one model and scenario, as this might modulate some of the key results of the paper (MC1). Key data underlying the paper should also be made publicly available before publication (MC3).

Thanks for a pleasant and informative read, and I’m looking forward to seeing future studies stemming out from ISA-MIP.

Moderate comments

MC1) The role of the Cerro Hudson eruption is really an important question. Checking out the latest version of the MSVOLSO2L4 inventory (curated by the NASA and Simon Carn, https://disc.gsfc.nasa.gov/datasets/MSVOLSO2L4_4/summary), the Hudson eruptions injected 4Tg SO_2 at 12-18km altitudes. The Neely and Schmidt (2016) inventory reports 1.5Tg SO_2 between 11 and 16km. So that would be between 7-40% of the Pinatubo mass depending on which value you consider for Cerro Hudson and for Pinatubo, a big number in any case. Could you repeat simulations, for at least one model and one of your scenarios, with Cerro Hudson included? I would actually suggest running one with the lower-end emission (Neely and Schmidt) and one with the upper end emission (MSVOLSO2L4). If you run only one set of parameters for Hudson, I strongly suggest picking a SO_2 mass in between these two estimates and not just the lower estimate (which is the one mentioned in your manuscript). Doing this test would really add a lot to the paper.

MC2) The injection strategy in UM-UKCA is really different. My personal experience with this model is that it's hard to get any SH transport unless injection is spread between 0 and 15N as done in your paper, and I think this is documented in published papers by Dhomse, Mann and co-authors. I recommend that you acknowledge this more explicitly in section 2, and that this model is singled out in a similar way to EMAC on all figures (e.g. on Figure 2 add a * symbol like you did for EMAC, and same everywhere else). It would be valuable to add comparison of point vs 0-15N injection for this model, either by running a point injection for one of your scenarios or by using already existing/published runs. In table 1, I would replace "band" for the injection region by "0N-15N, 120E". I believe that "band" injection would be understood by most people in this community as a zonal injection at the volcano latitude following the terminology used in e.g. Zanchettin et al. (2016) and Clyne et al. (2021), so "band" is misleading here.

MC3) I see no comment on data availability which is crucial before publication. In particular, having SI tables or a netcdf archive with the processed data displayed on key figures (for both model and observations) would be really welcome (at least for figures 2, 5, 8). This would facilitate comparison to your results for future studies.

MC4) This one is more a remark than a comment. This is a really nice paper and I strongly recommend prompt publication, but it's too bad that there aren't more modelling groups that ran the ISA-MIP HErSEA experiment in time for this paper or didn't follow the protocol. Out of the four models that followed the experimental protocol, three have some version of the ECHAM model at their core which limits model diversity especially when bias in circulation and subtropical barrier are suggested to be one of the main challenges. Figure 1 in Clyne et al. (2021) also suggest that the model used in this paper will produce middle-range SAOD estimates.

Is there no chance to include results from the IPSL or WACCM groups in this paper? Or to run the UM-UKCA simulations following the experimental set-up of figure 1 including point injection (or at least repeat the med scenario with the same 0-15N injection but a 19km height)? I realize that this is likely challenging at this stage especially for my first question. If so, my only recommendations are to acknowledge a bit more explicitly the lack of model diversity wrt the two points above (ECHAM as core model and middle-range SAOD estimates), and maybe to add a few sentences towards the end of the paper reflecting on what we can do as a community to encourage stronger participation to such MIPs? This could help the community leverage more funding and/or computing resources to support such intercomparison exercises.

Note: I realize that running additional simulations as suggested in MC1, MC2 and MC4 requires time and resources. However, the simulations suggested would use the same set-up as the ones already ran for the paper, so I hope that at least some of them are feasible within a reasonable timeframe given the atmosphere-only setup and small ensemble sizes/duration. The order of my comments reflects the priority I'd give to these additional simulations.

Minor and editorial comments

Line 3: Replace "plume" by "cloud" (here and throughout the paper). You mostly use "cloud" later, and "plume" is very commonly used for the vertically rising column rather than the large-scale horizontally (mostly) spreading cloud.

Line 17: The link with ash will not be obvious to a non-expert reader, could you contextualize briefly?

Line 18: add the country or latitude in parenthesis after "Cerro Hudson" so that the link is easier to make for non-expert readers.

Line 22: delete "can"

Line 29: "framework" instead of "frame"?

Line 30 and section 1: you have many paragraphs that are 3-5 line long; consider grouping some of them.

Lines 36-38: you could maybe point to earlier measurements and more recent papers to contextualize both the SO2 and ash injection height. Fero et al. (2009, https://doi.org/10.1016/j.jvolgeores.2009.03.011) seems particularly relevant. The IVESPA database (http://ivespa.co.uk/, endorsed by IAVCEI) also has best estimate and uncertainties based on extensive literature compilation for many events including Pinatubo. For Pinatubo the height of the plume top, ash injection height and SO2 injection height are 32+/-3 km asl, 22+/-3 km asl and 25+/-3 km asl.

Line 42: “are constrained across participating models”: do you mean that they are the same for all participating models right? I think the language could be a bit more clear.

Line 45: “This approach…has been shown to reduce discrepancies in reproducing …anomalies”. Compared to what other approach?

Line 39-55: Overall I find these paragraphs a bit hard to follow. Make sure that the language is explicit for the non-expert reader, and I would suggest reorganizing them a bit: i) start by describing results of the Tambora experiment and large discrepancies between models; then highlight consequences i.e. ii) the use of a single set of aerosol optical properties derived from a simplistic model for VolMIP; and iii) the need for ISA-MIP.

Line 54: Do you mean “lifetime” instead of “amount”? Sure different lifetime will ultimately affect the evolution of the aerosol burden, but lifetime would reflect better the characteristic affected by the effective radius.

Line 61: replace “initial conditions” by volcanic emission source parameters” or something like that to be more explicit?

Line 77: Why not also comparing the radiative forcing to observations? I guess this falls more under the remit of VolMIP, but it would still be of interest to many people to see which set of model/eruption source parameters result in the most realistic forcing? Radiative flux at the top of atmosphere are available from the ERBE instrument.

Line 85-87: Maybe briefly discuss what’s a realistic thickness for the injected SO2 cloud? I’m not sure if we have good constraints for the Pinatubo SO2 cloud. 3D plume model simulation suggest that the thickess of the gas phase should be about 10% of the column height (see Figure S2 in Aubry et al., 2019, https://doi.org/10.1029/2019GL083975).

Line 90: Explicitly acknowledge why SO2 is injected in this way in UM-UKCA, i.e. it’s already trying to fix the lack of SH transport in this model. This is a major difference in the injection set-up and UM-UKCA should be singled out on all figures/tables like EMAC (see MC2).

Line 91: For EMAC, either here or in the EMAC section, give more details on what these 3D-mixing ratio are in particular clarify how long after the eruption these 3D perturbation were constrained from observations (days? Weeks?), whether the injection date is modified accordingly in the model (it could affect e.g. the time at which peak SAOD is reached). Please also clarify the total mass of SO2 injected for Pinatubo and Hudson in

EMAC for comparison with other experiments.

Line 95: But I guess SO2 radiative effect (or ash) is not included in any of the models? It might be worth briefly acknowledging and discussing Stenchikov et al. (2021, https://doi.org/10.1029/2020JD033829)

Line 100: so only one ensemble member for ULAQ right? Make this explicit.

Section 2.1.1: You don't discuss at all the initial QBO phase. It looks like there was no attempt to pick a phase consistent with that at the time of the Pinatubo eruption (although models with nudged QBO will have this right, which isn't explicitly discussed)? This should be discussed for sure with citations of corresponding literature. How much would QBO phase affect your results in particular in terms of aerosol residence time in the tropics and SH transport?

Line 103: I would find it clearer if you replaced "six" by "five" and in the next sentence say something like "closely related simulations from a sixth model, EMAC, are considered".

Line 119-120: Maybe try to improve consistency in terms of the order of information given across model subsections? It will make comparison easier for the reader. E.g. always have horizontal and vertical resolution after the list of models coupled, then information on QBO, then information on microphysics, etc.

Line 121-122: you don't include information on how ensemble were produced for other model so be consistent? Also I'm not too familiar with this method. How long before 1991 was the rate of snow formation changed? I guess it would take some times to get really different initial states?

Line 136: Acknowledge somewhere explicitely that 4 models out of 6 have some version of ECHAM as their host model (also see MC4)

Line 174: Are these the same SST dataset as mentioned line 97? If so redundant info.

Line 177: I obviously know nothing about author contributions in the Schallock et al. (2021) paper, but I was surprised not to see the lead author of this study among the co-authors or mentioned in the acknowledgement section given the use of the Schallock et al. (2021) simulations.

Line 189: Here or where injection strategy for all models is discussed, give more details on these 3D injections.

Table 1: “Band” is misleading, see MC2

Section 2.2: Using the ERBE radiative flux and adding a figure comparing simulated vs observed TOA forcing would be a nice addition, even though this is more VolMIP than ISA-MIP remit

Lines 209-215 and 221-223: Could you clarify assumptions – e.g. on aerosol size distribution – required to derive parameters describing the aerosol (surface area density, effective radius, etc) from observations of optical properties? Should “observations” for these parameters be considered equal to e.g. SAOD observations or the direct balloon measurements?

Line 250: give resolution in degree latitude instead, and specify somewhere that GloSSAC provides zonally averaged values

Line 252: “tropical cloud core” instead of “tropical core”?

Line 255-256: I don’t find it that clear that Med-22 significantly overestimate SAOD for ULAQ, UKCA and EMAC?

Line 259: don’t use “band”

Lines 258-261: If the result that SH transport can’t be reproduced holds when including Cerro Hudson (MC1), you might want to formulate more explicitly the hypothesis that point injection is not a viable option for large-magnitude eruptions?

Line 261: at some point in the SH transport discussion (here or later in the paper), you might want to briefly mention Jones et al. (2017, https://www.nature.com/articles/s41467-017-01606-0), especially their figure 1? For the HadGEM model, it shows transport towards both hemispheres for a 23-28km injection but not for a 16-23km injection. This also motivates my comment MC2 to run point injection with UM-UKCA at different heights.

Line 276-277: true but the SH:NH SAOD ratio also looks pretty bad for this model?

Table 2 caption: be explicit about what correlation is considered here, and what RMSD, and also refer to appendix A1 for more details (same comment for figure 3 caption)

Table 2: add stars for EMAC and UM-UKCA here and in every figure/table. In captions you could say something like "* highlight models with spatially spread SO2 injections."

Line 284: really too bad that there is no experiment with other heights for UM-UKCA, nor experiment with point source (MC2). A few additional experiments would take a maximum of one or two weeks to run on UK HPC systems? Marshall et al. (2019, https://doi.org/10.1029/2018JD028675) should be discussed at some point for the role of injection height in UM-UKCA.

Figure 2 caption: Could you discuss briefly here and/or in the main text how big of a difference is expected between SAOD/extinction between the minimum and maximum wavelength used in different models/observational dataset? Checking Pinatubo simulations with the EVA_H model (an extension of Matt Toohey's EVA), I get up to 5% differences between 525nm and 600nm for global mean SAOD. I don't think the wavelength difference would affect your results (e.g. error metric, best scenario) too much but this should be acknowledged more clearly.

Figure 3: to make this figure easier to read, maybe you could have an empty taylor diagram at the bottom right of the figure with labelled arrows showing what metric changes how when moving one direction or another on the diagram.

Figure 4: Obviously important discrepancies between AVHRR and GloSSAC between month 8 and 21, but there is an apparent sudden "bump" around month 10. Could this be Cerro Hudson? (cf MC1) ECHAM6 and SOCOL capture very well the beginning and end of the AOD decrease.

Figure 4: add star for UM-UKCA; it would be nice to have the raw global mean SAOD values provided as supplementary data (also see MC3).

Line 286: I'm not a fan of using this definition to calculate the e-folding time as: i) it uses a single threshold instead of capturing the full decay trend in the data; ii) it uses the SAOD instead of the total S burden, and the SAOD is affected by things like the effective radius etc (it makes more sense to fit a mass decay than a SAOD decay). On point (i) could you quickly test if your results are comparable if you instead get the e-folding times by fitting exponential decay models to the data in Figure 4 (on a linear or log scale)?

Line 328: “This might depend on the different vertical concentrations of OH in the model”: be explicit on whether they increase or decrease with altitude and whether this is consistent with SO2 burden evolution.

Line 332-334: briefly discuss how consistent these results are with observational constraint on SO2 e-folding time dependence on altitude (see Figure 14 in Carn et al. 2016, http://dx.doi.org/10.1016/j.jvolgeores.2016.01.002)

Line 341: I’m not sure why this should be the case. Sure the characteristic timescale for SO2 -> sulfate aerosol conversion is shorter than the sulfate aerosol lifetime, but there will be a more or less small fraction (depending on injection height and mass) of sulfate aerosol lost before the full mass of SO2 is converted into aerosol?

Line 350: replace “by” by “with”

Line 351: Here and everywhere else where you say “injection rate”, replace by “injected SO2 mass”. The key parameter is how much SO2 you inject, not how quickly you inject it in the models (even though this might also have an influence especially when comparing basaltic to silicic eruptions, but it’s not the aim of your experimental design)

Line 352: “Figure 3 shows that the differences” (that instead of comma)

Line 354: do you mean 22km instead of 19 for the three scenarios?

Line 365-367: please see MC1 and update the range of plausible eruption source parameters to 0.75-2Tg S and 12-18km with citation of MSVOLSO2L4 and Neely and Schmidt (2016, https://doi.org/10.5285/76ebdc0b-0eed-4f70-b89e-55e606bcd568). In IVESPA (see earlier comment), for the largest phase of the Cerro Hudson eruption, we have 16+/-3km for the plume top height and 17.5+/-3km for the ash injection height, with no good constraint found for the SO2 height.

Line 369: peak location of what?

Line 383: you mean panel b and e instead of c and f?

Line 386: “injection rate” -> correct everywhere, see previous comment

Line 390: does not instead of doesn’t

Line 391: remove one occurrence of “especially …after the eruption”

Line 390-391: Acknowledge Marshall et al. (2019) where they show that higher injection heights result in aerosol being in slower branch of the BDC and longer tropical confinement?

Line 396: “in which aerosols…high latitudes” -> mention that this effect is season-dependent?

Line 411: How is the mean effective radius calculated? Is it weighted by e.g. aerosol concentration? If not you might get large differences purely related to the vertical distribution of aerosols in the different datasets?

Line 418: “steady” instead of “flat”?

Figure 7: replace “ratio” by “aerosol mass fraction”?.

Figure 7 g-i: Is the sum of each row not equal to 100% because of aerosol outside 60S-60N? This really confuses me. If so could you standardize wrt the mass within 60S-60N?

Figure 7: Why is the +/-10% band highlighted in grey? Is this deemed a reasonable agreement and if so how do you justify the threshold? If no justification just have a horizontal line at 0 instead.

Figure 7 caption: the burden (mass) is an extensive variable so it makes no sense to take its spatial average. Do you mean “total burden” instead of “global average burden”?

Line 426: add “of ISA-MIP” after “experiment”.

Line 429: “since the simulated decay onset time is anticipated”: I don’t understand what this means, reformulate please.

Line 456-457: This refers to figure 9c? The discrepancy between observations is much smaller than the inter-model spread though?

Line 493: replace “mechanism” by “process”?

Line 501-503: comment on how UKCA differ? While noting that the injection strategy differ.

Line 514: could or might be crucial, not would?

Line 513: in addition to a longer lifetime it would result to slower latitudinal transport because BDC speed decreases with height? Also cite Stenchikov et al. (2021) in this paragraph.

Line 520: At least one experiment with Cerro Hudson (MC1) would be really good to test how the lifetime is sensitive to the inclusion of this additional eruption.

Line 525-528: this sentence is very long and hard to follow; please rephrase and break down.

Line 531: define w* for the non-expert reader

Line 534-535: not much discussion on that, and in particular you barely discuss QBO configuration in your experiments?

Line 535: another relevant reference is Jones et al (2016, https://doi.org/10.1002/2016JD025001)

Line 540: Do you mean 18-25km

Line 563: Also cite the recent perspective paper by Marshall et al. (2022, https://doi.org/10.1007/s00445-022-01559-3)