Authors' response to reviewers' comments

Reviewer 1

This study focuses on the global and regional (tropical, subtropical, and springtime polar) impacts of a new SAI experiment on stratospheric ozone changes and potential drivers. This experiment mainly differs from previous GeoMIP experiments by injecting SO_2 at symmetric subtropical latitudes instead of equatorial bands, and using a pathway of lower GHGs, namely SSP2-4.5 compared to SSP5-8.5, and reducing the warming to roughly present-day temperatures. It applies a multi-model approach to differentiate between inter-model and SAI experiment differences. Chemical and dynamical effects are separated using additional model runs that exclude heterogeneous chemistry on SAI aerosols. The simulated SAI experiments exhibits smaller lower stratospheric warming compared to equatorial injection scenarios, which translates into smaller stratospheric circulation changes.

The studies' main findings include the need for more accurate modelling of stratospheric transport, as inter-model differences appear to be driven by these. Generally, inter-model differences dominate over differences within injection scenarios in the one tested model. There seems to be an asymmetry between the polar responses. The authors attribute the decrease in column ozone to halogen catalysed chemical loss, and transport changes induced by SAI. As process importance differs between different latitudinal regions, the authors separate different contributions and attribute dynamical response differences to process level uncertainty in stratospheric transport.

The below-mentioned comments are not critical, but require some changes that go beyond simple re-wording. Thus, I suggest moderate revisions of the manuscript as outlined below.

We thank the reviewer for their review. We address the individual points below in blue.

Since heterogeneous halogen activation on sulfate aerosols is put forward as one of the major factors in decreasing ozone columns, a brief introduction of the mechanism, corresponding studies, and model differences would be helpful.

We agree but note we already introduce the different chemical, dynamical and radiative processes by which SAI can affect ozone in the introduction (paragraph two), give details of the heterogeneous reactions in Section 2 and Table S1, and then discuss the different mechanisms in detail throughout the manuscript.

At multiple instances in the manuscript, the authors refer to either "heterogenous" or "heterogeneous" chemistry. Please check for consistency.

Thank you for spotting this – we have corrected all instances of heterogenous to heterogeneous.

At multiple instances, the authors write TO3 or O3, please correct this to O_3.

We have searched the manuscript but we only use "TO3" and "O_3" in the figure labels and captions, and not main text. At any rate, these two abbreviations mean two different things, with TO3 denoting total column ozone and O_3 denoting ozone. We now define these terms when they first appear in the figure captions.

Please check for consistency when referring to sulfate and sulfate aerosols.

We believe it is common for these two terms to be used interchangeably, and since at different instances it is helpful to have either longer or shorter term (for clarity or avoiding repetition, respectively) we prefer to keep both terms in the text.

l.26: Instead of, or additionally to, "more plausible middle-of-the-road", please mention the SSP.

Added.

l.29: "compared to [a] no-SAI"

Now corrected.

l.35: While agree with the strategy dimension, I don't recall analysis of the scenario dimension. At least in case the latter refers to different SSPs. To keep this argument, some further analysis to isolate its influence should be included, which may already be present in previous G6-1.5K-SAI publications.

The cited sentence is a general statement on how our results fit into the wider SAI research space and how the results motivate future work; it is not meant to indicate that we include analysis of scenario dimension in our study.

l.38: A reference for ozone's shielding role, and effects on life and ecosystems should be included here.

This statement is a common knowledge statement, and as such does not require a citation: https://usingsources.fas.harvard.edu/exception-common-knowledge.

l.43: Since the impacts of climate intervention methods are highly uncertain, a more careful formulation, such as "to [potentially] temporarily offset", should be considered here.

We have rephrased that to "… proposed methods ***aiming*** to temporarily offset ..."

l.52: This full sentence lacks a reference to acknowledge previous work.

References to this are already given in the opening sentence of that paragraph: "SAI can affect stratospheric ozone via a range of chemical and dynamical processes (e.g. Pittari, et al., 2014; Haywood and Tilmes 2022; Bednarz et al., 2023a,b)."

From the APA style guide: "Overcitation can be distracting and is unnecessary. For example, it is considered overcitation to repeat the same citation in every sentence when the source and topic have not changed. Instead, when paraphrasing a key point in more than one sentence within a paragraph, cite the source in the first sentence in which it is relevant and do not repeat the citation in subsequent sentences as long as the source remains clear and unchanged."

https://apastyle.apa.org/style-grammar-guidelines/citations/appropriate-citation

ll.53-55: This statement also lacks at least one, or several references.

References to this are given in the opening sentence of that paragraph: "SAI can affect stratospheric ozone via a range of chemical and dynamical processes (e.g. Pittari, et al., 2014; Haywood and Tilmes 2022; Bednarz et al., 2023a,b)." (see also comment above) ll.64-66: This statement lacks a reference to heterogeneous chemistry of aerosols. At this stage, more details on halogen chemistry, and results from other models (e.g., SOCOL, MPI-ESM) with aerosol chemistry could be added.

Heterogenous chemistry on aerosols does not require a reference; it's a known process (https://usingsources.fas.harvard.edu/exception-common-knowledge).

We have now added a sentence about the previous results from the multi-model CCMI-2022 senD2sai experiment: "Most recently, Jörimann et al. (2026) used a multi-model ensemble of SAI simulations participating in the ongoing Phase 2 of the Chemistry Climate Model Intercomparison project (CCMI-2022; Plummer et al. 2021). While providing invaluable insights on the models' sensitivities to imposed stratospheric sulfate aerosols in an atmosphere-only set-up, the experiment, by design, excluded some important sources of uncertainty driving the aerosol, climate and ozone responses to the intervention."

l.68: Please reference the GeoMIP project properly here.

GeoMIP is already referenced at the end of the sentence (using Kravitz et al. 2015 as a reference).

l.77: Consider hyphenating "middle-of-the-road".

This is already hyphenated.

l.79: Consider hyphenating "business-as-usual".

This is already hyphenated.

l.100: "each of the model[s']"

Now corrected.

Figure 1 (a): This figure would benefit from a shading or similar representation of errors to better estimate model differences. This is especially the case for the runs with SAI. The authors may also consider a visualisation and brief discussion of the inter-model differences in the chosen base-line states, and their impacts on their findings.

A visualization and discussion of intra-ensemble spread and variability (including differences between individual ensemble members) and inter-model differences in the chosen baseline state is outside of the scope of the current manuscript (which instead evaluates simulated ozone responses), and is instead analyzed and discussed in detail in Lee et al. (2025, https://doi.org/10.5194/egusphere-2025-5742, see their Fig. 1 and Section 3.1)

We have added a statement of this to the manuscript: "The analysis of intra-ensemble variability in the simulated global mean near-surface temperatures, including differences in baseline temperature targets, is found in Lee et al. (2025)."

l.115: Please clearly state the CESM version.

We already clearly state the CESM version - "the Community Earth System Model version 2 coupled to the Whole Atmosphere Community Climate Model version 6, CESM2(WACCM6)" l.121: "including [the] most"

Now corrected.

Table S1: Please write "2 Cl", "2 Br", ...; please also restructure the table's title, for instance moving "in each model" to the end of the title.

We have changed this as requested.

ll.160-162: This statement regarding the coarse mode may also deserve a citation.

We have added a reference to Sekiya et al., 2016.

ll.162-164: This statement regarding the SAD would pose as significantly increased uncertainty of chemical responses and would deem SAD as a much weaker constraint for this response. It would potentially also limit the extent to which chemical responses in UKESM imply responses in MIROC. This aspect deserves some further discussion in the corresponding section (section 7 or 8).

The uncertainty in the simulated SAD response is just one of many sources of uncertainty in the diagnosed ozone responses to SAI. We have added this statement to the text. However, we're not sure what the reviewer means by "the extent to which chemical responses in UKESM imply responses in MIROC". In general, a response in model A does not imply the same response in model B, as models in general differ in their representation of various processes.

Figure 2: A visualisation of the uncertainties would be useful. Subfigure (c) uses a different period (2050-2069) on the right side compared to the left side. These differences should be discussed and evaluated. In the same subfigure, the relation of ozone and surface temperature change seems to be ambiguous and without clear physical connection. In case the authors want to use this metric, it should be more thoroughly introduced before.

Regarding the visualization of uncertainties in Fig. 2, we already include the measure of the corresponding uncertainties in panel C. We think adding intra-ensemble spread also to panels A and B would make the figure harder to read. In addition, we would like to keep the composition of Figs. 2, 3, 8 and 10 consistent throughout. In the reviewer's comment on Figure 3 below, the reviewer points out that panel 3a is very busy as it is (i.e. without a measure of intra-ensemble spread) and suggests dropping ensemble mean data altogether and only plotting running means. While here for Figure 2, i.e. one that is analogous to Fig. 3a, the reviewer suggests adding extra lines. Since we want to keep composition of figures consistent, we prefer to stick to plotting only the ensemble mean and its running mean data in panels a and b (in Figs, 2,3,8,10), and then including a measure of intra-ensemble and inter-annual uncertainty as error bars in panels c.

Regarding the different period used in the analysis of the simulations of Bednarz et al. (2023a) and G6-1.5K-SAI and the need to discuss that – we already discuss and evaluate this on line 224-226 (preprint version): "For the strategy injecting SO2 at 30N+30S, the ozone response from the simulations in Bednarz et al. (2023a), diagnosed over 2050-2069, falls well within the range of ozone responses diagnosed from the G6-1.5K-SAI simulations over 2045-2064 and 2065-2084."

Regarding the last point, we disagree with the reviewer that the relation between ozone and surface temperature change is ambiguous and without clear physical connection. Both large-scale surface cooling and stratospheric ozone changes are directly driven by sulfate aerosols from SAI (with global surface cooling being the main objective of SAI). We believe it is thus appropriate to evaluate the ozone response per degree of the associated surface cooling.

l.180: lacks a reference to BDC changes following GHG emissions.

We have added a reference to WMO (2022).

ll.184-186: The statement on dynamical effects on ozone lacks references, being itself not straightforward, nor directly related to the results of this study.

The role of dynamical effects on ozone in the lower stratosphere is now well established but as requested we have added a citation to WMO (2022) for this. Since our study is on stratospheric ozone projections under SAI compared to no-SAI scenario, we believe thar giving context on ozone changes occurring under no-SAI scenario is important and directly related to this study.

ll.186-188: The connection between global-mean total column ozone and surface temperature is ambiguous, and neglects the complex relationship between the two. It appears that the authors aim to connect the ozone column changes directly to potentially policy-relevant changes in surface temperature. By doing so, the underlying processes are blurred and physical understanding is lost. To my knowledge, no direct connection between these two quantities exists that would allow to link them in this way. I recommend to either exclude this metric (ozone column change per degree of warming), or justify its use and physical foundation more clearly. This oversimplification must be removed in the final version of the manuscript. This holds for all plots using this metric.

We disagree with the reviewer that the relation between ozone and surface temperature change is ambiguous and without clear physical connection. Both large-scale surface cooling and stratospheric ozone changes are directly driven by sulfate aerosols from SAI (with global surface cooling being the main objective of SAI). It is ***not*** that column ozone changes are expected to drive significant changes in surface temperatures, rather than SAI drives changes in both ozone and temperature. We believe it is thus appropriate to evaluate the ozone response per degree of the associated surface cooling.

l.192: "larger earlier in the century than later" --> This doesn't really hold for CESM, except from the mid-2070s.

We do not agree. As shown by the red line in Fig. 2b, the relative ozone reduction in CESM is indeed larger earlier in the century than later.

l.194: "progressively higher" --> the aerosol burden curve seems to flatten, whereas temperature difference increases roughly linearly. This may be discussed in the context of separating the earlier to later parts of the century, instead of mainly focussing on decreasing halogen concentration.

Both the ***difference*** in surface temperatures between each model's SSP2-4.5 and SAI simulations (i.e. the difference between dashed and solid lines in Fig. 1a) and the SO2 injections in the SAI runs (Fig. 1b) show varying rate of increase between the early period (approx. 0 to 10 years after SAI start) and later on. This is because initially temperatures in the SAI simulations are a bit higher than the target and, hence, the initial SO2 injections need to not only offset the rising temperatures under rising GHG emissions (which as the reviewer points out increases quasi-monotonically) but also provide some extra cooling to bring the temperatures down to the target levels. After the first decade or so, i.e. once global temperatures are roughly at their target values, the rate of change of both relative difference in TAS between SSP2-4.5 and SAI simulations and the SO2 injection rates change slightly. At any rate, this should not affect the discussion regarding differences between 2045-2064 and 2065-2084 periods, as both of these periods analyzed occur after the initial ~10-year period.

ll.224-226: I don't agree with this statement for the earlier period from 2045-2064. Please clarify this confusion, and also rather refer to the figure than, or in addition to, the original paper to clarify the comparison here.

We have rephrased "falls within the range of" "falls in between the range of". What we meant here is that the 2050-2069 response falls in between the responses diagnosed for 2045-2064 and 2065-2084 (so within the range given by two decadal periods on either side of 2050-2069). We have also added a reference to Fig. 2c earlier on in this paragraph.

l.229: Please reference the G6sulfur experiment.

The reference for G6sulfur (Kravitz et al. 2015) is already given earlier on in the introduction, but we have now added a reference to Lee et al. 2025 that compares some of the large scale surface climate changes between G6sulfur and G6-1.5K-SAI.

ll.229-230: "and thus potentially less favourable outcome" --> I suggest to omit this part, being speculative and not adding information.

We have changed this to "This highlights the trade-offs between optimizing surface climate response and minimizing negative impacts on stratospheric ozone".

Figure 3: (a) is very busy, and may justify only showing the running mean (with uncertainty); (b) shows a jump (with change in sign) for CESM which may be discussed in more detail to ensure it being physical.

(a) As discussed above regarding the earlier reviewer's comment concerning Fig. 2, we would like to keep the composition of Figs. 2, 3, 8 and 10 consistent throughout. For Fig. 2a, the reviewer suggested adding extra lines. Here the reviewer suggests removing some of the existing ones. We believe the sensible compromise here is to keep the figures as is, i.e. show both ensemble mean and running mean data in panels a-b, then include a measure on intra-ensemble and intra-annual variability in panels c.

(b) as the reviewer points out in (a), the forced response to either GHG or SAI is small and the internal variability is large, which can manifest as the interannual variability in the diagnosed response.

Figure 3 (c) uses 'DU' as the y-axis as physical quantity, whereas it should be 'TO_3' in [DU]. Similarly, Figure 6 confuses physical quantities and units. Please check all Figures regarding these.

We have corrected this to "[DU per 1 K]" in Fig. 2c, Fig. 3c, Fig. 8c,f, Fig. 9, Fig. 10c,f, as well as changed y-axes labels in Figs. 6 and 11 accordingly.

Figure 4 lacks indications of the physical quantity (pressure) and its units (hPa) on the y-axis; please also add the diagnosed tropopause, since it is referred to at several instances in the text, but seems to be defined inconsistently.

Thank you for spotting this – we have added "[hPa]" as y-axis label. We prefer not to add the approximate tropopause to the plot as to avoid the confusion between the tropopause and the other contour lines on the plot (which instead indicate climatological values in SSP2-4.5).

ll.250-251: I only agree with the decrease in tropical ozone column from around 2050 in CESM and MIROC, and from around 2070 in UKESM.

We have changed this to "Unlike the projected increase in global mean total column ozone, the tropical ozone column is projected to decrease slightly under the SSP2-4.5 scenario over the second part of the 21st century (Keeble et al., 2021), with differences in the projected trends, including their signs, across different modes (Davis et al, 2026; also Fig. 3a for the models used here)."

l.252: Please reference the figure showing the change in the BDC here.

The relevant citation for this (Keeble at al., 2017) is given already at the end of that sentence.

ll.252-254: The connection between lower stratospheric decrease in ozone and upper stratospheric ozone is not clear. Please reformulate (and split) this sentence to clarify the argument.

We have rephrased and clarified this sentence: "This arises because of offsetting between the ozone decrease in the lower stratosphere and ozone increase in the upper stratosphere (Fig. S2a-c). The lower stratospheric ozone reduction arises because of the GHG-induced acceleration of the BDC and the resulting increased transport of ozone-poor tropospheric air into the lower stratosphere; the ozone increase in the upper stratosphere arises due to reduction in chemical ozone loss as halogen levels are reduced and the stratosphere cools (e.g. Keeble et al., 2017; Davis et al., 2026). "

l.256: The difference in sign of the ozone response appears like an offset. A discussion of this effect between the models would be useful here. For instance, adding a factor to the UKESM response could shift it towards the observed CESM values. Within this framework, a brief discussion on the influence of the early phases of SAI on the following trajectory or sensitivity tests could be insightful, and further make the results more robust.

As discussed in the paper (including Section 4 in question), different chemical and dynamical processes determine the resulting ozone response to SAI in different regions, and different processes vary differently with time. Heterogeneous halogen activation is most important in the early period and then its importance is the combined result of decreasing background halogen levels available for activation and increasing sulfate concentrations. On the other hand, N2O5 hydrolysis and all dynamical changes tend to increase with time due to increasing sulfate concentrations. Model differences in all of these complex processes lead to differences in magnitude of ozone changes across the models and the rate of such changes with time. We do not believe adding a factor to the UKESM response would be appropriate, and performing sensitivity tests to isolate the drivers of the early phase ozone response to SAI is beyond the scope of this study.

ll.257-258: MIROC seems to drift more significantly towards more negative values in Figure 3 (a). A short discussion of possible reasons and corresponding uncertainty would be useful.

We have now added in the text that models generally disagree with respect to their projected tropical ozone trends (including adding relevant references). The more negative tropical ozone trend in MIROC is just a manifestation of this general model uncertainty. The discussion of potential reasons behind the diverging tropical ozone trends under rising GHGs in different models is beyond the scope of the current manuscript, and the reviewer is encouraged to refer to other manuscripts investigating this topic (e.g. Davis et al. 2026, https://doi.org/10.5194/egusphere-2026-532) l.267, l.276, l.280: The mentioning of the lower stratosphere and tropopause seems to use different definitions of these regions. Please ensure to consistently refer to the same regions, and annotate them within the corresponding figures.

We don't see why the reviewer thinks we use different definitions of the lower stratosphere, these terms are used consistently across the manuscript.

ll.272-274: This statement requires a reference, or more elaborate evaluation.

The statement that total column ozone levels in the tropics are lower and insolation is higher is a fact, and as such does not require a reference.

l. 283: Please be more concrete in linking stratospheric warming to circulation changes. For instance, one may rather refer to "stratospheric circulation" and its specific features (e.g., BDC) rather than circulation in general.

We have specified this and changed this to "stratospheric circulation"

ll.283-285: This statement is not new, and deserves acknowledgement of previous work.

Previous work that showed that is already acknowledged in the preceding sentence (Bednarz et al., 2023a,b). We have now added a third reference (Henry et al., 2024) to that sentence.

ll.304-307: "The much lower climatological stratospheric water vapor in MIROC (Fig. 7h) might reduce aerosol water uptake and hence partially contribute to the model showing comparable amount of heterogenous halogen activation and halogen-catalyzed ozone depletion than the other two models (Section 3, also Sections 5-6), despite much larger aerosol SAD (Section 2, Fig. 1d)." --> Aerosol water uptake, or hygroscopic growth, mainly depends on Relative Humidity (RH) and not stratospheric water vapour. The former determines the thermodynamic equilibrium vapour pressure over binary solutions such as H_2SO_4-H_2O droplets. RH is generally very low in the stratosphere, resulting in large decrease in uptake sensitivity with height. The climatological absolute humidity in MIROC may be significantly lower than in the other two models, but so may saturation vapour pressure, resulting in relatively constant RH, and thus aerosol water uptake. I suggest to check RH for differences between MIROC and the other two models to constrain this statement.

We have rephrased this to say: "The much lower climatological stratospheric water vapor in MIROC (Fig. 7h) might act to reduce relative humidity and hence aerosol water uptake, thus partially contribute to …"

Figure 5: In case no AoA figure is added for CESM, please change the naming of the subfigures to (a)-(e). This figure also deserves an annotated diagnosed tropopause.

We have corrected figure panel labels. We prefer not to add the approximate tropopause to the plot as to avoid the confusion between the tropopause and the other contour lines on the plot (which instead indicate climatological AOA values in SSP2-4.5).

Figure 6: Please write (a) and (b) in the caption, not (A) and (B). Write-out the abbreviation "TEM" in full please.

We have corrected this and defined both TEM and w* terms in the figure caption.

ll. 342-344: Increased transport by the BDC resulting in increased ozone lacks a reference.

We have added a reference to WMO (2022).

Figure 8 would perhaps benefit from using relative instead of absolute units for ozone columns. Such a visualisation may also limit the model dependence on inter-model differences in climatology.

Multi-model projections of total column ozone and its changes are commonly shown using absolute units (i.e. DU), both in previous published results from multi-model intercomparison (e.g. Keeble et al., 2021; Dhomse et al., 2018) as well as in WMO/UNEP Ozone Assessment reports. Since we would like our results to be comparable with previous works, we prefer to keep showing absolute units (DU) in our plots.

l. 350: I would agree with this statement for CESM and MIROC, but not for CESM.

Even in the UKESM the NH and SH mid-latitude ozone response to SAI is different, both in long-term trend and interannual variability.

ll. 352-354: Please provide some reference or figure for this statement. In this context, some justification for the presence (and absence) of these difference may be added.

We have added a reference to Fig. 7(a-c). The justification for these differences is already given at the beginning of next paragraph ("A very different picture emerges for the NH mid-latitude ozone (Fig. 8a-c). Here, higher climatological lower stratospheric temperatures than in the SH (Fig. 5) slow down the rates of heterogeneous halogen activation on sulfate, and so any SAI-induced changes in transport can be as important in determining the overall impact on ozone.").

ll.363-364: "higher climatological" than what?

We have added "than in the SH"

ll.382-384: I agree that figures 9 and 8c illustrate the change in TO3, but don't see how they allow to attribute these to transport changes.

We have added a reference to Fig. 6b and Bednarz et al. (2023a).

ll.384-385: Please show, or reference, the difference in the importance of chemistry vs. transport in a figure or similar to support this statement.

We have rephrased this to "Second, the models appear to disagree with respect to the importance of heterogeneous chemistry versus transport for the NH mid-latitude ozone response. Only MIROC projects a significantly more negative normalized ozone response in the early (2045-2064) versus late (2065-2084) period (Fig. 8c and 9). This suggests that heterogeneous processes are not of dominant importance for determining the NH mid-latitude ozone response to SAI in CESM or UKESM, which instead likely favor more SAI-induced transport changes and dynamical impacts on ozone (which tend to scale with the amount of surface warming and stratospheric heating, Fig. 6)"

l.394: Figure 5d is non-existent.

We have corrected figure panel labels

ll.410-411: The projected increase in ozone is still debated and not as clear as stated here.

We disagree. This is now well established and supported by previous research studies, numerous inter-model comparison projects (e.g. Keeble et al. 2021, https://doi.org/10.5194/acp-21-5015-2021; Dhomse et al. 2018, https://doi.org/10.5194/acp-18-8409-2018; Eyring et al., 2010, https://doi.org/10.5194/acp-10-9451-2010; Eyring et al. 2013, https://doi.org/10.1002/jgrd.50316) and recurring international WMO/UNEP Ozone Assessment Reports (e.g. 2014, 2018, 2022), all of which show that springtime Antarctic ozone is projected to increase significantly over the 21st century under a range of emission scenarios, and the magnitude of the total column ozone increase there is projected to be the largest from the other latitudinal regions. We have now added reference to Keeble at al. (2021) and Dhomse et al. (2018) to that sentence to further back it up.

l.413: Just write CESM and MIROC here, without the later negation.

No, this is true also for UKESM but only between 2035 and mid-2070s. We have rephrased the text in parenthesis to make this clearer ("though UKESM starts to show a small increase relative to SSP2-4.5 around the mid 2070s")

ll.416-419: This statement blurs climate sensitivity and ozone columns, and deserves justification or correction.

We don't see the problem with this statement. We first clearly state the absolute ozone changes, and then in the second statement give the ozone changes normalized with the associated SAI-induced surface cooling. By definition, the normalized ozone values reflect additional uncertainty from differences in model climate sensitivities (both to rising GHG and sulfate aerosols from SAI), and the main point of this metric is to incorporate this uncertainty.

l.442: "Domaisen" -> "Domeisen"

Now corrected.

ll.453-454: Please reference the corresponding figure here (10b).

Now added.

ll.473-475: This statement deserves a reference.

This is a general statement that does not require a reference.

ll.478-480: Neglecting differences in surface cooling and injection rate and deeming them less relevant deserves further justification.

As we state in that sentence, these two simulations have "***slightly*** different injection rates and surface cooling values". Any slight differences are, by definition, not expected to lead to first order differences in simulated responses. In addition, as indicated in the sentence, Section 2 discusses and justifies this statement in depth: "Nonetheless, we expect this overcooling to have only a marginal impact on stratospheric ozone, especially when analyzing the differences between the SAI and no-SAI simulations, the effect of which is expected to be by far offset by the effect of including heterogeneous chemistry on SAI sulfate. In addition, we present many of the results as values normalized with the magnitude of the associated global mean near-surface SAI-induced cooling, hence further removing effects from small inconsistencies in the temperature target definitions."

ll.486-488: This statement lacks justification on why it has to be halogen catalyzed loss, and not other chemical processes, or chemistry-dynamics interactions.

The only other heterogeneous chemical process occurring on sulfate is N2O5 hydrolysis, which acts to increase ozone levels in the middle stratosphere and hence cannot explain ozone reductions in the lower stratosphere. And we aimed to state "partly by the halogen catalyzed chemical ozone loss" as to indicate that such ozone loss is coupled with circulation response, but the word "partly" had a typo (as per comment below), so we now removed this word and explicitly state "halogen catalyzed ozone loss and its coupling with circulation (Section 6)".

l.487: "pearly" -> "clearly"

This was meant to say "partly" but given the comment above, we have now removed this word and explicitly state "and its coupling with circulation (Section 6)".

ll.545-548: The direct comparison to Tilmes+2022 seems difficult since it uses a different "strategy" and scenario. I agree that mentioning these differences is relevant, but directly inferring conclusions from them obscures the differences and complexity.

We have added ", although differences in GHG and SAI scenario could be a further potential contributing factor"

l.579: I agree with the importance of mentioning the comparison alongside SSP2-4.5. However, its implications for the analysis should be further elaborated to justify its mentioning.

What we tried to say here is "Importantly, the SAI impacts discussed in this paper are largely shown as changes relative to the SSP2-4.5 simulation over the same time period" (we have now rephrased this), i.e. these are largely changes calculated as SAI minus SSP2-4.5 over the same time. The remainder of the paragraph is the elaboration of implications of this for the analysis/conclusions.

l.581: The "lagging" of the ozone response is not strictly shown here. Especially from 2060, the ozone responses rather seem to converge. This statement requires further justification.

We have now changed this to ", just lagging the ozone evolution projected under no-SAI scenario, at least until near the end of the 21st century"

ll.583-584: Please provide reference(s) for this "common assumption".

We have rephrased this part.

ll.584-586: I don't see the relevance or direct connection to the results and objectives of this study.

The fact that ozone levels are supposed to 'super-recover' under global warming scenarios (including SSP2-4.5 used here) compared to historical values, the implications of which for human health and ecosystems are still not fully understood but may be both positive or negative, have direct implications for whether the impacts of any potential future SAI-induced ozone decrease relative to a no-SAI case would be positive or negative. We have rephrased this to make this clearer.

Reviewer 2

Paper overview

The manuscript presents work on estimating the stratospheric ozone impacts from the stratospheric aerosol injection (SAI) model intercomparison project GeoMIP, experiment G6-1.5K-SAI. The study shows results from three different Earth system models for future scenario modelling using the SSP2-4.5 scenario. The results on ozone impacts are discussed for two different future time periods and the analysis is split into different regions of the Earth as well as focused on the global impacts on ozone.

I think the manuscript provides a usual set of results for the scientific community and new information on the ozone impacts from SAI, necessary before widespread use of these suggested technologies. However, I have several points that need to be addressed before the manuscript is ready for publication, with one point being the clarity of writing. The presentation of the results lacks clarity in the main messages, which would enhance the importance of the results further. I recommend major revisions as described below in the overall comments followed by line-by-line comments.

We thank the reviewer for their review. We address the individual points below in blue.

Overall major comments

Only MIROC includes an ensemble of 10 members, UKESM and CESM have 3 or 5 members, which is a very low number to obtain robust climate varying temperature scenarios. The temperature changes are central in the analysis of the resulting ozone reductions in the future scenarios as presented in the paper. Please comment on this and address the issue in the introduction and discussion of the results.

A discussion of intra-ensemble spread and variability, including differences between individual ensemble members in their resulting global mean surface temperature changes, is outside of the scope of the current manuscript (which instead evaluates simulated ozone responses), and is instead analyzed and discussed in detail in Lee et al. (2025, https://doi.org/10.5194/egusphere-2025-5742, see their Fig. 1 and Section 3.1). We have now added a statement of this to the manuscript: "The analysis of intra-ensemble variability in the simulated global mean near-surface temperatures, including differences in baseline temperature targets, is found in Lee et al. (2025)."

However, we note that while 3 or 5 (or even 10) ensemble members may be insufficient to confidently diagnose many regional surface temperature responses, the variability in global mean annual mean TAS (which is most important for the experimental design, and not regional) is much lower and hence much easier to constrain, as confirmed by the TAS values successfully converging around their target levels in the three models.

The UKESM model runs with heterogeneous chemistry only on background or SAI aerosols make the comparison between the three models difficult as they're not compared like for like. Also, the three models have chemical mechanisms for halogen heterogeneous chemistry that are significantly different, which needs to be highlighted in the beginning of the paper as well as in the discussion of the results, as this could potentially be a main driver of the observed model differences.

As discussed in Section 2, all three models compared include heterogeneous chemistry on sulfate aerosols from SAI, which is the main chemical driver of the diagnosed SAI responses. Whether or not a model also includes heterogeneous chemistry on background sulfate (the concentrations of which are significantly lower than the sulfate from SAI) is highly unlikely to influence the diagnosed ozone response to SAI, as this is always calculated relative to the control simulation with exactly the same background aerosol treatment as in the perturbed simulation (and so any effect this might have on ozone cancels out). Moreover, the UKESM simulation without heterogeneous chemistry on SAI sulfate is only used in Section 7 as a comparison against the UKESM simulation with it, and not against the other two models. Again, whether each simulation pair includes heterogenous chemistry on background sulfate will not affect the diagnosed response to SAI (calculated as perturbed minus control) as any effect of this will cancel out.

Regarding the role differences in heterogenous chemical mechanisms, we state and document these differences in Section 2 (and Table S1). We believe it would be incorrect to highlight these differences as the main driver of the inter-model differences in the diagnosed ozone response to SAI. The final ozone response is determined by a range of different processes, including a similarly important contribution from dynamical and transport processes, all of which are subject to similar uncertainties and differences in model representation of the various physical processes at play. In addition, these differences in the complexity of heterogenous chemical mechanisms are not straightforwardly evident in the diagnosed ozone responses (for instance, CESM includes much smaller number of heterogeneous chemical reactions on sulfate than MIROC but it shows the largest Antarctic and SH mid-latitude ozone loss from SAI).

What would this analysis show if you focused on stratospheric ozone and not the total column? Are there any differences between the models that would be exaggerated or diminished? How much do these injections influence tropospheric ozone concentrations? i.e. through stratosphere to troposphere transport? Discussion of this would benefit the study.

A significant driver of the tropospheric ozone response to SAI is the change in lower stratospheric ozone concentrations and the associated stratosphere-to-troposphere ozone transport (e.g. Xia et al., 2017, https://doi.org/10.5194/acp-17-11913-2017; Tilmes et al., 2022, https://doi.org/10.5194/acp-22-4557-2022; Wang et al., 2026, https://doi.org/10.5194/acp-26-1339-2026), as it is likely the case here in the SH in CESM and MIROC. We note that vertical profiles of ozone responses are included in the paper (not just total column), and differences in ozone at different pressure levels are discussed. However, isolating the tropospheric ozone column responses and diagnosing their drivers in these simulations is beyond the scope of this manuscript, but should instead be focused on in future follow up studies.

The writing is very convoluted in places, with a lot of inserted sentences in the main sentence. This makes it difficult to ascertain the key messages, and the findings of the study are obscured. Please go through the text and check the longer sentences for clarity and possible ambiguity. Examples are highlighted in the line-by-line comments.

We have checked the manuscript in light of this.

The figures need to be more streamlined and described in full. I have several points for you to go through the figures in the main text and in the supplementary material: 1) All figures need to be described precisely and in detail and discussed in the main text or moved to the supplementary material. 2) Each figure must be able to stand alone, to be assessed without previous figures or additional publications. 3) In some figures and parts of figures you refer to/label figures as annual mean data, where you are presenting multi-year averages. Please be precise in how you refer to this and avoid labelling things as annual means if it is not that. 4) Include a secondary vertical axis of altitudes in units of km to all vertical data as well as a line highlighting the tropopause height (if this changes in the SAI vs SSP model run, choose which one and describe this in the caption/text where needed). 5) Add tick marks on the right-hand side of your figures to make it easier for the reader to analyze the figures. 6) The y-axis is plotted vs. the x-axis, not the other way around. Please change this in how you refer to the data plotted on the figures.

Thank you for these points. We address the individual points in the line-by-line comments below.

Line-by-line comments

Line 50: "… that impact concentrations of species directly relevant for chemical ozone production and loss." This should be specified in more detail. It reads as overly vague as it is, the details would be helpful to have here in the introduction as you discuss the specific species in the results.

This paragraph introduces the range of chemical, dynamical and radiative processes that affect ozone under SAI. Line 50 dealing with heterogeneous chemistry already states that this refers to halogen and nitrogen reactions. A detailed list of reactions in question is given in Section 2 and Table S1, and the role these reactions play for ozone is discussed throughout the manuscript. We think that adding a lot more details to this paragraph on various heterogenous chemistry reactions would make it disproportionally focused on this one particular aspect, while the main point of the entire paragraph is to introduce the range of different processes that matter for ozone under SAI.

Line 53: "Any local changes…" are these in addition to and/or separate from the aerosol-induced temperature changes? In the line above you discuss temperature changes from additional aerosol, this sentence is not clear if this is from the same aerosol change, please clarify. I think you mean to say something more generally about impacts of changing temperatures in the stratosphere than only from sulfate aerosols in lines 53-55, however, this needs to be clearer.

We have rephrased this to "Any resulting changes ..". This is meant to encompass both direct aerosol-induced temperature changes from aerosol absorption as well as any resulting indirect temperature changes from modulation of dynamics and wave mean flow coupling (e.g. adiabatic cooling/warming).

Line 85: "30N+30S" – missing degree signs, please add them and replace the plus-sign with "and". Adding the note here to remind you to go over the manuscript and check for inconsistencies in how you refer to injection locations.

Now corrected.

Line 98-101: Two comments here: 1) Figure 1b is mentioned before 1a, this would suggest that you should switch the order of those two panels in your figure, so they're labelled in the order they appear in the text. 2) I'm unsure how the injections were done reading this sentence. Did you use the resulting surface temperature to adjust the amount of SO2 to inject? The baseline that you are adjusting to include years where SO2 is injected – why was this chosen as the reference point? Can you please add a bit more information and re-phrase to argue the way the injection has been set up?

(1) We now mention Figure 1a earlier in that paragraph first.

(2) As already stated in the text (i) we indeed use the resulting surface temperature to adjust the amount of SO2 to inject: "with the magnitude of the injection (Figure 1b) adjusted interactively in each model at the beginning of a year in order to maintain the global mean near-surface air temperatures at the baseline level …", and (ii) no, the baseline we're adjusting to doesn't include years where SO2 is injected as it's defined using the SSP2-4.5 simulation: "… at the baseline level, taken as the ensemble mean 2020-2039 average of each of the model's SSP2-4.5 ensemble"

Figure 1: How can you show an annual mean "YM" for panels c, d, and e? I think you are showing a mean of all the years 2045-2064. Please remove the YM label on the panels and see comments on editing the figure caption below. Formatting: Almost all text on the figure is too small, the labels on the colorbars are too small to read and you're missing degree signs, subscripts on units as well as showing mixed styles of panels in the sense of font types and sizes. For this figure you only need one colorbar for the bottom row of pnaels as you show three identical colorbars here. Please check this and the other figures in the manuscript to make sure they are legible and have correct units.

While the shading shows a multi annual (2045-2064) mean, we are using annual means to calculate this, which becomes important when calculating statistical significance. A statistical significance test calculated based on, for instance, monthly mean or daily mean data within the 2045-2064 range would be different to statistical significance calculated using annual mean (YM) values. As such it is important to include the word annual mean (or abbreviation YM) in the current description as it contains important information.

We have now improved the figure formatting and added the missing superscript and now have only one color bar.

Figure 1 caption: Please add in the caption text exactly what is shown in each panel. It is important to note the model names in the caption. Line 105: "annual mean global mean", please rephrase to the more commonly used "global annual mean" or similar. Bottom row: "annual mean 2045-2064 changes", so you are showing changes between the two years or the mean of all those years? Or did you take the mean of each year and then do something else? In line 109-110, you say "between G6-1.5K-SAI and SSP2-4.5 over the same period in each model". Reading this I conclude that you have taken a mean of 2045-2064 (not "annual" as these are multiple years) for two model ensembles, see comment above about "YM": 1) SSP-reference run and 2) SAI model run. The "same time period" refers to what exactly? I'm not convinced you should have a range of years to begin with and then also add this. This caption needs to be clarified.

We disagree with the reviewer regarding the need to add model names to the already long figure captions, given that this is self-explanatory and described in depth throughout the manuscript.

Panels in bottom row are calculated by first calculating annual mean values for each year in the 2045-2064 range, then calculating a mean over 2045-2064 for each ensemble and taking the difference. While not influencing the mean 2045-2064 value (so shading on the plot), the fact that the figure is calculated based on annual mean data is important for correctly calculating and interpreting statistical significance, hence we believe it's important to keep it here. We have rephrased this to "Annual 2045-2064 mean" for clarity.

Line 115: Here you refer to the model as "CESM2(WACCM)". This results in two brackets next to each other; this should be avoided by rephrasing. I.e., this could be achieved by the common way of referring to the model "CESM2/WACCM" using a slash instead of bracket.

We are unaware of the model name being referred to using a slash, CESM2(WACCM6) or CESM2-WACCM6 are the two standard ways of doing so in previous literature.

Line 116: Whis has the "ES2H" part of MIROS abbreviation not been described? You describe version numbers for the other two models, leaving it out here is inconsistent.

ES2H is not an acronym. See the reference for the model (Kawamiya et al. 2020).

Line 132-145: I am not following how these four runs in total relate to each other from this text. The caption in SI is a lot cleaner and easier to follow. Can you please rephrase this paragraph to resemble more the Si fig caption? The mention of original simulations (from the Lee-study) and discussing how different simulations are used instead is confusing – it is not clear why you would necessarily want to repeat the published work by Lee et al. How are you using their temperature finding here if you are using a different set of model simulations?

We have rephrased this paragraph to make it clearer. The main point is that the Lee et al. (2025) study is just describing the default, publicly-available G6-1.5K-SAI simulations that most SAI analysis on has, or will, use. But since the original UKESM simulation that constitutes the official G6-1.5K-SAI experiment (as used in Lee et al.) is not suitable for our study (because of the absence of heterogeneous chemistry on SAI sulfate) we here describe a revised UKESM simulation we use in the main part of this study instead.

Line 142: "our revised simulations" which simulations are you referring to here?

We have rephrased and clarified this paragraph.

Line 142-146: "the same absolute target global mean near surface temperature from the original SSP2-4.5 experiment was used in the revised G6-1.5K-SAI simulations, leading to a slight overcooling (~0.2 K) compared to the reference 2020-2039 mean surface temperature in the revised SSP2-4.5 runs (Fig. 1a). We note this slight overcooling occurs under smaller SO_2 injection rates (Fig. S1c), as the SAI-induced ozone reduction (Section 3) acts to amplify the aerosol-induced surface cooling." From the referred figures, I don't see this overcooling being very significant until the end of the century – if this is what you want to highlight, say that this is the timeframe for comparison. However, why did you not adjust the SO_2 injected to account for the differences between the NO_HET and WITH_HET runs? And is the difference in SO_2 injections in S1c really significant? I see the resulting difference in S1a, but related to that: please explain why the difference between the two SSP runs in S1a diverge more as time passes?

Yes, this overcooling isn't very significant, and this is why we describe it as "slight overcooling" in the sentence in question. And while this could have been avoided by updating the temperature target that controls the SO_2 injections, we mistakenly didn't do that, and re-running the simulations to correct this small inconsistency would not justify the large computational resources this would entail. And so we use the simulations as they are now, just making sure this feature - and its relevance to the results (if any) - is properly described in the text. We are convinced that such an approach is appropriate and justified.

Figure 2: I am concerned regarding panel a that the model differences are largely due to the different chemical mechanisms as described in the supplementary material as UKESM and CESM are showing the largest increases in column ozone, particularly UKESM is showing recovery of global column ozone over time in panel a, and not the GHG emissions differences as you highlight in line 186-188. Can you add a description/discussion on why this is not the case, please. And/or evaluate/assess the reasons for the differences from different chemistry vs different GHG emissions? How are you determining the uncertainties represented as error bars? I am missing a description of this in the text. Finally, same as for the previous figure: YM shouldn't be used in panel c as this is a multi-year mean not the annual mean of one year.

We are confused by what the reviewer is asking. The line 196-188 in question states: "The inter-model differences in the magnitude of the projected total column ozone increase are qualitatively consistent with the associated differences in the GHG-induced global warming, with UKESM showing the largest concurrent global mean near-surface temperature increase and MIROC the smallest (Fig. 1a)." It does not state that these differences are driven by the GHG emission differences, as the reviewer implies. In fact, these differences cannot be driven by GHG emission differences as all models, by experiment design, have exactly the same GHG emissions, as given by the SSP2-4.5 scenario. Rather, this sentence implies that these ozone differences are consistent with inter-model differences in climate sensitivity.

Secondly, the reviewer points to Table S1 which discusses differences in heterogeneous chemical mechanisms occurring in the models on sulfate aerosols only. However, long-term changes in stratospheric ozone over the 20th and 21st century are primarily driven by long-term changes in long-lived halogenated ozone depleting substances (ODSs) and GHGs; the former modules chemical ozone loss by a combination of gas-phase and heterogeneous reactions, primarily on Polar Stratospheric Clouds (PSCs), while the latter affects ozone by changing stratospheric circulation and ozone transport, and by modulating chemical ozone loss via changes in stratospheric temperatures. While some halogen catalyzed ozone loss is also accelerated by heterogenous reactions on background sulfate aerosols, or resulting liquid binary solutions, especially in the Arctic region (e.g. Solomon et al. 2015; https://doi.org/10.1002/2015JD023365), this effect is only one of the number of processes affecting long-term evolution of stratospheric ozone. We do not see how evaluation and assessment of drivers of inter-model differences in the projected long-term changes in ozone under conventional global warming scenario (i.e. without SAI), including the relative role of chemical vs dynamical drivers of the diagnosed differences, lies within the scope of our paper, which instead focuses on the evaluation and assessment on inter-model differences in the ozone responses to SAI. The reviewer is encouraged to refer to previous publications on this topic (e.g. Dhomse et al. 2018; Keeble et al., 2021).

Regarding the calculation of uncertainties, the caption to the figure in question (Figure 2) includes how the uncertainty range is calculated: "The error bars represent +/- 2 standard errors in the difference in means, calculated from the annual mean values (20 per ensemble member)." And as per the response to the previous reviewer's comment, the mention and use of annual means is central here for correctly calculating and interpreting the statistical significance test, hence we believe this should not be removed.

Figure 2 caption: line 173: CESM2(WACCM-MA) is not described before here, and abbreviation is not spelled out in the main text. Please add this information.

CESM2(WACCM-MA) is described in the main text (Section 3). We have now spelled out the abbreviation in the main text.

Line 183-184: "As shown in Figure S2(a-c), while models generally agree on the magnitude of upper stratospheric ozone increase" the way this is phrased it sounds like you are showing a model comparison of the upper stratosphere, which cannot be seen in figure 2, please rephrase to clarify what you are showing in the figure and what is a separate statement.

This is phrased correctly. The sentence is referring to Figure S2, not Figure 2, and this figure shows vertical ozone profile responses (including upper stratospheric ozone response)

Line 192-194: "The latter occurs despite the increasing SO2 injection rates - and hence increasing stratospheric aerosol burden - over the course of the 21st century in order to offset progressively higher surface temperatures under rising GHGs in SSP2-4.5 (Fig. 1a-b)." The meaning of this sentence gets lost with the inserted sentence referring to aerosols burden, can you rephrase this to make it clear that you are increasing the injected SO2 over time to account for increasing temperatures due to increasing GHGs in the SSP scenario, please. Highlighting that what you adjusting the models to achieve is the deltaT/deltaSO2 as described in your table 1.

What is important for ozone is not increasing SO2 injection per se but increasing stratospheric aerosol burden. Hence the inserted sentence is an important part here. We have changed the dashes to a parenthesis, for ease of reading.

Line 203: all these numbers seem very detailed; do you really have enough confidence to claim two significant digits for the DU ozone changes? Please check the statistical significance of these numbers. Also, these changes are all very small – what are they equivalent to in percentage? You refer to percentages earlier in this paragraph, please make these numbers easily comparable to that.

We have updated the values in this sentence to: "we find global mean ozone column losses of -0.4 DU/Tg-SO2·yr-1 in 2045-2064 and -0.15 to -0.3 DU/Tg-SO2·yr-1 in 2065-2084". The reason why we think it's better to keep the second number exceptionally with two decimal places here is that when rounded to one decimal place (0.1455 ≈ 0.1) this make it incorrectly indicate there is a factor of 3 difference between the models (i.e. 0.1 vs 0.3), when the actual difference is a factor of 2 (i.e. 0.1455 vs 0.2935).

We cannot simply add a percentage value to be comparable to the percentage value given earlier, as the exact number would depend on how the percentage is defined. Previously, the % value referred to absolute ozone changes, and the percentage was defined compared to the baseline (2020-2039). Here the ozone values are the relative ozone changes per degree cooling; the actual baseline periods are also different in the simulations of Bednarz 2023a, as in those simulations global mean surface temperatures are controlled to achieve 1.0 K above preindustrial levels (corresponding to the baseline of 2007-2028 mean), hence defining the % change in relation to the 2020-2039 mean would be misleading.

Line 208: The years in the bracket refers to the literature data from Bednarz et al 2023a, the bracket is misplaced here as it suggests that you are showing new data in this time range. Please make it clear that you are comparing the present study's two averages to the literature 2050-2069 values for the different strategies. Maybe talking the years out of the bracket and rephrasing to accommodate this action or moving the bracketed years to a later place in the sentence (i.e., line 209) would work to clarify this along with adding the time ranges for the present study. Cf. your writing in line 226.

We have now moved this bracket to later on in that sentence to make this clear.

Line 213: please add altitudes and/or pressure levels for your definition of "middle atmosphere"

Middle atmosphere is a term commonly used to generically refer to the part of the atmosphere above the troposphere and below the thermosphere and ionosphere, see e.g. the textbook by Andrews et al. (1987). In the sentence in question, it refers to the name of the chemistry scheme.

Andrews, D. G., Holton, J. R., and Leovy, C. B.: Middle Atmosphere Dynamics, Academic Press, San Diego, 489 pp., 1987.

Line 221: please add altitude and/or pressure levels for your definition of "middle-stratosphere", also remove the hyphen to be consistent with the previous text, see comment on line 213. Highlighting this here to bring to your attention to check the manuscript for this and similar inconsistencies.

We have removed the hyphen. The sentence summarizes the published findings of Bednarz et al. (2023a), where no clear-cut range of pressure levels definition of the middle stratosphere is given. We now make this clear in the text. At any rate, we disagree that referring to processes occurring in a region of the atmosphere denoted generally as the 'middle stratosphere' is an inconsistency, but a deliberate choice given that, and highlighting that, those processes do not stop occurring at a pressure level X or Y hPa, and so giving finite pressure level range would be factually incorrect.

Line 222 and 223: “heterogenous N2O5 hydrolysis on sulfate” and “halogen activation on sulfate”, I think you mean to say “sulfate aerosols” here, please correct this.

It is common for the terms “sulfate aerosols” and “sulfate” to be used interchangeably.

Line 224: 1) “lower stratosphere”, see previous comments, add altitude to your definition. 2) “at 30N+30S” this is not a location when written this way, but a model simulation label. Please be precise and consistent in how you refer to the model simulations, in this case it should be “in 30N+30S” to refer to a model simulation or replace with the proper latitudes with degree signs.

1. See the response to the comment above.
2. We have now changed this as suggested.

Figure 3: Please add a proper caption for this figure It needs to be able to standalone from figure 2. Panel a has a legend that covers some of the data, please correct this to allow inspection of that missing data in the ensemble means. Scale the graph to cover all of your data, missing the yellow UKESM peak around 2058 in both panels a and b and maybe also for 2035 in panel b

We disagree with the reviewer that each figure needs to be able to stand alone regarding its captions, as opposed to referring to previous figures if most of the information is the same; instead we believe this is more the matter of specific journal requirements and/or personal preference. We have however updated the figure caption as requested.

We have moved the legend in panel a to top right as requested. We disagree with the reviewer’s suggestion regarding panel b. The current scale allows for a proper inspection of the models’ long-term responses to SAI. Scaling the plot per reviewer’s suggestion would make these forced long-term changes more difficult to see and intercompare, while shifting the focus to parts of the plots which are just a reflection of natural variability and, hence, do not provide any additional information to the scientific question examined.

Figure 4 caption, line 246: remove “over the same period for each model” you are already saying the this is the changes for your two time periods in line 245, repeating it after mentioning the SSP scenario is confusing and unnecessary.

The reason we emphasized it this way is because it is a common approach in many SAI studies to compare changes in a SAI simulation in a future period against a control simulation in a present-day period. Since our analysis is deliberately different to that approach (as analyzing ozone responses to SAI when compared against the same period of a control simulation allows to better isolate the SAI response from ozone changes caused by other factors, e.g. long-term changes in ODS or GHGs), we think it’s important to make this clear which time period is used for the control simulation to avoid confusion.

Line 250-251: “the tropical ozone column is projected to decrease slightly under the SSP2-4.5 scenario over the 21st century (Fig. 3a)” I disagree on this part of the sentence: Inspecting figure 3a MIROC and CESM start decreasing after ~2050 and UKESM only after ~2070 if taken as all values are below the zero-line in your figure.

We have changed this to “Unlike the projected increase in global mean total column ozone, the tropical ozone column is projected to decrease slightly under the SSP2-4.5 scenario over the second part of the 21st century (Keeble et al., 2021), with differences in the projected trends, including their signs, across different modes (Davis et al, 2026; also Fig. 3a for the models used here).”

Line 251-252: “This arises because of the GHG-induced acceleration of the BDC increasing transport of ozone poor tropospheric air into the lower stratosphere” this statement needs a reference, please add this here.

The reference for this (Keeble et al., 2017) is already given at the end of this sentence.

Line 252-253: “the resulting lower stratospheric ozone reduction offsets the ozone increase in the upper stratosphere (Fig. S2a-c)” I do not see the upper stratospheric increase in the S2 figure, please expand your argument for this statement. Also, reading your vertical figures would be easier if you include information on the tropopause height in all vertical figures.

We are confused by this comment. Figure S2(a-c) clearly shows ozone increase in the upper stratosphere.

Line 265-266: “These differences in tropical total column ozone responses amongst the three models reflect the complex pattern of tropical ozone changes at different altitudes.” I think you need to add more information to this statement, as you are not including the complexity of the chemical schemes being different in each of the models as highlighted in your own Table S1. You need to add this complexity to the discussion, already here in the first sentence of this discussion or in the beginning of the paragraph. Your current chemistry discussion start in line 287.

This opening sentence, as well as the whole paragraph in question, discusses the simulated ozone changes. The drivers of the simulated ozone changes are discussed in detail in the following paragraphs. We believe it is important to first tell the reader what the simulated ozone responses are, and only then add more information on the drivers of the changes and any related information and complexity.

Line 273: typo “insolation”, should be “insulation”

No, the word “insulation” refers to “the act of covering something to / material that is used to stop heat, sound, or electricity from escaping or entering” according to Cambridge Dictionary, which has little meaning in an atmospheric context; we mean “insolation”, referring to the exposure to solar radiation.

Line 307-309: add “vapor” in a few places in this paragraph to make it clear that you are talking about the gas phase.

We have changed “H2O” everywhere to “water vapor”.

Line 320: figure panel labels: even when CESM doesn’t have a panel, the bottom to panels should still be d and e, not skipping the letter d. Please update this accordingly in the figure caption and main text.

We have corrected figure panel labels

Line 330-331: In your figure caption: “(B): As in (A) but for changes in the shallow branch of the BDC, defined as in Bednarz et al. (2023a), i.e. 15^{o}S-15^{o}N mean TEM w* changes at 100 hPa.” This text needs to explain what is shown in panel B, referring to Bednarz et al 2023a is necessary, but this figure and the paper should be able to stand alone. Therefore, you should add more details to this caption, i.e., explain the definition of the shallow BDC branch, “TEM” abbreviation, and describe “w*”

We have rephrased this to “… 15^{o}S-15^{o}N mean of the Transformed Eulerian Mean (TEM) residual vertical velocity (w*) changes at 100 hPa”

Line 335: Add “vapor” to the mention of water or add “gas-phase” somewhere in the first line to describe all three species. I am missing the precision in the descriptions of what the figures show, see overall comment as well.

We have changed H2O to water vapor.

Figure 8: All panels, but in particular panels c and f need to be explained in the figure caption. It should be stressed that some panels are showing changes in DU and c and f in DU/deg K

As it was for Figure 2, we disagree that each figure needs to necessarily be able to stand alone regarding its captions, as opposed to referring to previous figures if most of the information is the same. We have however updated the figure caption as requested.

Line 350-361: discusses figure 8 panels e-f, the following paragraph discusses figure 8 a-c. Please move and re-phrase the text to discuss the panels in order (or re-organize the figure panels to match the order they are mentioned). This will help the reader follow the logic of the text and figure together better.

We have re-ordered panels a-c and d-f in Fig. 8.

Line 351-352: You write: “… reductions of up to approx. -10 DU (with the 2045-2064 average of -7 to -9 DU SH mid-latitude ozone loss depending on the model) under SAI compared to SSP2-4.5 throughout the length of the simulation (Fig. 8e).” However, in panel 8e, I see values as low as ~-11 DU for UKESM and CESM in that time range (values below -10 DU). The panel is very difficult to read accurately, making evaluating the numbers difficult, however, please update these numbers in your range to match the data.

A value of -11 DU is “approx. -10 DU”. The numbers in the text are correct. We now specify we talk about “***long-term*** ozone reductions”, not short-term deviations caused by natural variability.

Line 363-364: Your text: “Here, higher climatological lower stratospheric temperatures (Fig. 5) slow down the rates of heterogenous halogen activation on sulfate…”. Can you please add discussion on reactions that increase with lower temperatures? The ClONO2 hydrolysis reaction is a key example of where the reaction becomes less important at higher temperatures, which is the opposite to your overall statement. This information needs to be included in the discussion if you want to give a statement about halogen chemistry.

We are confused by the reviewer’s comment. As the reviewer correctly states, “the ClONO2 hydrolysis reaction is a key example of where the reaction becomes less important at higher temperatures” (in fact, all heterogeneous halogen reactions on sulfate slow down under higher temperatures – as we imply in the text - as they are a function of collision frequency, which in turn is an inverse function of temperature). So we do not follow why this is opposite to our overall statement that “higher climatological lower stratospheric temperatures (Fig. 5) slow down the rates of heterogenous halogen activation on sulfate…”.

Line 370-371: In your text: “These differences are likely a manifestation of large differences amongst the models in their projected chemical and dynamical responses to SAI,” please add the differences in halogen chemical mechanisms as an additional explanation

The difference in “their projected chemical [..] responses” encompasses all the different drivers of those differences (i.e. including halogen chemical mechanisms).

Line 385-388: Please comment here on the fact the MIROC has the most complete heterogeneous chemistry of all three models and the impact that has on the following discussions of transport vs. chemistry importance in MIROC and the other two models. This is necessary and has not been addressed sufficiently in the current manuscript, see overall comments.

We already point out that MIROC includes the largest number of heterogeneous chemical reactions on sulfate in Section 2: "All models include a number of most important heterogeneous chlorine, bromine and nitrogen reactions, with MIROC including the most comprehensive list of reactions." However, we don't have the required evidence to isolate this particular aspect as the dominant driver of some of the differences in the simulated ozone responses (as discussed in response to previous comments).

Line 409-410 and following paragraphs: Here you start with discussing the SH referring to panels 10d-f, please re-organize the figure or the text so the panels appear in the order they are discussed, starting with panel a.

We have re-ordered panels a-c and d-f in Fig. 10.

Line 418: "…at the same time…" please add the years to make it clearer that you refer to 2045-2064 averages

We think this is clear as is.

Line 419-421: please add what the ozone loss ranges are for the later period to contrast the 2045-2064 averages with the 2065-2084 ones.

The 2065-2084 normalized ozone loss values are -22 DU/K in CESM, -14 DU/K in MIROC and 0 DU/K in UKESM, and the corresponding ozone losses normalized with the associated SO2 injections rates ranging between -2.4 to 0 DU/Tg-SO_2 yr^{-1}. We have now added this.

Line 422-425: please comment on the large difference between the CESM 2045-2064 average in the main panel of figure 10f and the past CESM2 run with the 30S+30N strategy. In this figure they look very different, which hasn't been the case for the previous figures of the same setup.

We are not sure of the exact reason. The 2050-2069 ozone response from the previous study is still in between the range of ozone responses diagnosed here for 2045-2064 and 2065-2084, but the reviewer correctly points out that it is much different from the 2045-2064 response despite more similar time period used. We note that the older CESM2 simulations and the G6-1.5K-SAI runs were carried out using two different WACCM versions, the former tailored for middle atmosphere studies and the latter including detailed representation of chemistry from the troposphere to lower thermosphere. While these should technically include the same representation of stratospheric chemistry, it is possible that factors other than that – e.g. tropospheric aerosol emissions – influence Antarctic ozone response indirectly, e.g. by affecting background tropospheric and stratospheric climatology and variability.

Figure 11: 1) Clarify what the deltas refer to for all panels in this figure – I assume you mean (SAI – SSP2-4.5), but this should be explicit in the figure, text, and caption. 2) panels c and d are delta O3 vs. the delta U, not as the titles say, 3) why does panel d include MAM for O3 and DJFMAM for delta U? This discrepancy is not described as it should.

1) Deltas in Figure 11 are calculated in exactly the same way (i.e. SAI – SSP2-4.5 over the same period) as in all other previous similar panels. Moreover, the caption states "As in Figure 6 but for …", and the fact this is shown as SAI - SSP2-4.5 is reiterated clearly in Figure 6 caption.

2) We now corrected the titles in c and d.

3) DJFMAM was chosen instead of MAM as it includes information about the polar vortex strength (and hence indirectly temperature) during most of the dynamically active season that is relevant for heterogenous halogen activation and determining springtime (MAM) Arctic ozone. This is unlike in the Antarctic where changes in winter (JJA) polar vortex are not as important for springtime (SON) Antarctic ozone, as JJA polar vortex is almost always strong enough to allow PSCs formation and heterogenous halogen activation, and so the variability in the springtime polar ozone is driven primarily by the polar vortex variability in spring. The choice of averaging period is clearly stated in both the main text and figure caption, and hence we do not agree it is an inconsistency.

Line 453: please add reference to figure 10 to help the reader understand what figure you are describing – I see no reference to figure 10 panels a-c in the text; this needs to be rectified.

We have added a reference to Fig. 10a-c.

Line 453-462: In this paragraph, please add more description of figure 10a-c and of the relationship described in figure 11b and 11d. The two panels in figure 11 for the NH are referenced in the paragraph, but more description and discussion of the results are needed for background of the conclusion drawn in the paragraph. Please rephrase the text accordingly.

The text already contains a clear reference to the figure, statement of what is seen in the figure, statement of the main conclusion drawn from the figure, as well as discussion of the result and relation to previous work (see extract below). We are unsure what else could we add here.

"Regarding the relationship with the concurrent winter-to-springtime (December-to-May, DJFMAM) polar vortex changes, while all three models show overall strengthening of the Arctic vortex under SAI (Fig. 11b), no obvious relationship is found with the springtime Arctic ozone loss across the three models and two time periods (Fig. 11d). In general, while changes in Arctic ozone have been found to be connected to changes in Arctic polar vortex strength under different contexts (Friedel, 2022a,b; Chiodo et al. 2023, Kult-Herdin et al., 2022), such correlations tend to be more difficult to detect, and this could be partly because of much larger variability of the NH polar vortex, both in time and space, thus making such a relationship more difficult to diagnose using monthly and zonal mean data. "

Line 480: "…dominant differences will be driven by the presence or absence of heterogenous chemistry on SAI aerosol…" Even if you refer to section 2, you need to make a short statement of the main differences of heterogeneous chemistry in the two runs – the one used in the other parts of the paper doesn't have sulfate heterogeneous chemistry, but only on SAI, whereas both MIROC and CESM have heterogeneous reactions occurring on both types of particles at the same time. The caveats about using the two versions of UKESM along with the other two models and using the UKESM only to analyze chemistry vs. transport impacts need to be addressed in the discussion. See overall comment as well.

The main difference between the two runs here is the presence or absence of heterogeneous chemistry on sulfate aerosols from SAI, and this is clearly stated in the text. As discussed extensively in Section 2, the presence or absence of heterogenous chemistry on background sulfate aerosols is strongly unlikely to affect the diagnosed response to SAI, as this is always calculated against the response in the control simulation with exactly the same treatment of heterogeneous reactions on background (and so any impact of that cancels out).

If any, slight differences here might be driven by the slightly different injection rates and surface cooling values, and this is something we remind the reader in this paragraph. Finally, this section (section 7) does not mention the results from CESM and MIROC at all, so we are unsure why should we bring that up here.

Line 487: "…driven pearly by…" is this a typo? I think you mean to say "partly" or "clearly". Please fix this so the meaning becomes clear.

It was meant to say 'partly', but this word has now been deleted in the final version.

Line 491-493: these three lines are the only mention of figure 13, if you want to include it in the main text, please describe the information and analysis of it. At the moment it feels like it could be a supplementary figure.

The figure is described in detail in its caption, and we think the current language in the main text in is sufficient to convey the main message from this figure, which we think warrants inclusion in the main text.

Line 492: "…here approximated as altitudes above 12 km…" Please add the approximate pressure level that this refers to and add altitudes in km as a secondary y-axis on your vertical figures to aid the inspection of the data, see overall comment.

We have converted Fig. 13b to show total column ozone summed for altitudes of 150 hPa and above. This way all plots in the paper consistently use pressure as vertical coordinate, removing the need for a secondary y-axis with a crude approximation of altitude based on hydrostatic equation.

Line 498: This statement is not backed up with the data as it stands: "…ozone poor-air from the lowermost stratosphere (as also suggested by the corresponding changes in model age-of-air pointing towards younger air in these regions, Fig. S6)…" The main feature seen in figure S6 in all four panels is an increase in age-of-air in the polar regions below 300 hPa and in the tropics around 200 hPa and above. Below these altitudes you have this younger air (seen as a decrease). At around 30 hPa, some of the panels show decreases in age-of-air, but if this is what you mean to highlight it is not clear from the text. Please re-phrase this to match your data. The increases cannot be overlooked in your analysis, even if you want to focus on the decreases in age-of-air.

We have now rephrased this statement to make it clearer which regions we are talking about.

Note that we discuss AOA increases in the lower stratosphere in detail in Section 4 in relation to Fig. 5 and tropical ozone changes across the models – but here (i.e. Section 7) this lower stratospheric AOA increase is similar between the two UKESM simulations and hence will not contribute to the differences in the ozone responses between the two (and hence is not discussed in Section 7, where we chose to focus only on the aspects that can contribute to these differences).

Figure 13: caption: please take the "here taken as altitudes above 12 km" out of the panel b label and add the information as text in the caption, including it with the "(b)" hides the panel designation. Include which pressure levels (approximately) the >12 km refers to in the caption as well.

We are convinced it is important to keep the information on the vertical levels that are summed over in panel (b), otherwise the reader would be left confused as to why the two panels have exactly the same title but different results.

Line 512-521: Please delete this paragraph as it is repeating points already made in the introduction with no new added information. Replace by one sentence introducing the summary and conclusion section instead, or simply start the section from line 523.

We have reduced the text and reorganized it.

Line 528: "…increased heterogenous halogen activation on sulfate and the resulting enhancement of the halogen catalyzed ozone loss…" When you highlight this, you need to discuss the fact that the three models have very different halogen chemistry included, which complicates the comparison and analysis of the three model outputs. Somewhere close to this sentence, I would suggest to add that information.

As discussed above, different earth system and climate models in general differ in their representation of a range of chemical, microphysical, dynamical and radiative processes, and we believe it would be misleading to single out a particular aspect as the dominant driver of the differences in the simulated responses, especially without any evidence that it actually is the dominant contributor to those divergent ozone responses.

Line 541: wrong use of "albeit", it is synonymous to "though", "while", "but", etc., using it here gives the idea

that it is "in contrast to", which suggests that the models don't agree on the global ozone reductions, which I don't think is what you intended. You mean to say that the regional SH mid-latitude ozone loss is larger than the global ozone loss, correct? Please remove the brackets, delete "albeit" and add the comparison to global ozone reductions as part of the main text. Even if you just delete "albeit" and keep the rest of the sentence, it will improve the sentence and understanding for the reader.

Thank you for spotting this – we have removed the word 'albeit'.

Line 554-555: "…becomes a factor of ~2 smaller than earlier in the century…" Again, can you please add the numbers in the 2065-2084 range? Are the models still showing a spread of the same magnitude or do they converge? Is CESM still showing the largest ozone reduction? The numbers to remind us of the figures would be helpful.

The corresponding range of ozone loss for 2065-2084 is between 22 DU/K and 0 DU/K. We had deliberately chosen not to put these exact numbers in the text, as the results can be seen already in Figure 10 (and now have also been added to the main text in Section 6) and we think having too many numbers in the text, especially in the summary/discussion section, distracts the reader from the most important massage, i.e. that these 2065-2084 ozone losses are approximately a factor of 2 smaller than the 2045-2064 losses. Hence we prefer to keep the text here as is.

Line 566-577: You should discuss here the bias from using the UKESM chemistry versus using the more explicit one in MIROC and how this could potentially underestimate the importance of chemistry driving the changes. Particularly in relation to the discussion below of the model versus observations, where you solely point to transport changes. How would the discussion change if you add the complexity of varying the halogen heterogeneous chemical mechanisms in the models? Please comment and add this to the discussion in this paragraph.

We are unaware of any "bias" from using the UKESM chemistry. Similar to CESM, the UKESM model includes some of the most important heterogeneous chemical reactions on sulfate aerosols from SAI, making it fully suitable for studying SAI impacts on ozone. The UKESM version without such heterogeneous chemistry on SAI sulfate is used in the paper only as a sensitivity study to be compared against the default UKESM version with such processes included, and not against the results of other models.

Furthermore, it is not possible to evaluate the importance of varying complexity of heterogeneous halogen scheme in such coupled simulations, especially since the different models differ generally in their representation of a range of chemical, microphysical, dynamical and radiative processes, all of which contribute to the differences in the diagnosed ozone responses. If one assumed that the degree of complexity of heterogenous scheme is the dominant driver of the diagnosed differences in ozone responses, then MIROC - which has the most detailed representation of heterogeneous halogen reactions on sulfate (as the reviewer points out) as well as a factor of ~2 larger sulfate SAD changes (Fig. 1e) - should have the largest ozone loss, but this is not the case, as diagnosed ozone losses in the model were generally smaller than, or in between, those in CESM and UKESM.

Finally, we do not only point to the transport changes in this paragraph, but we also talk about how these changes “make an important contribution to the subtropical and mid-latitude ozone losses in the model”. It is stated clearly earlier on that heterogenous halogen chemistry constitutes an important (and in some regions dominant role) in driving the simulated ozone changed to SAI in the G6-1.5K-SAI simulations, and in this paragraph we focus on highlighting the previously largely unrecognized uncertainties in the transport changes and their contribution to ozone responses in the subtropics and mid-latitudes. (Also, the discussion of “model versus observations” centers on recent ozone trends, and not SAI, for which no observations exist).

Figure S3: add x-axis label to panel a, the ticks are missing degree signs and something to say “latitude” or similar for the axis. Please add a description for the terms mentioned in the figure. I.e., “ΔTAS”, “YM”, even when these are described in the main text.

We have now labelled x-axis as latitude. We note that ‘TAS’ and ‘YM’ were already previously defined in the supplement in Figure S1.