REVIEW #1

I don't know what to recommend for this paper. Clearly there was a lot of analysis of different models, and they were all different. But so what? What is the new science? What do we know now that we did not know before? What is the scientific question that is being addressed? Weren't these results found in previous papers, such as part of VolMIP? It would be very useful to know which models or models are actually correct.

You write that we should not depend on the results from just one model, so what should we do? Which model or models should we use? There have been injections into the stratosphere from volcanic eruptions and forest fires, for which we have observations. Have any of these models been used to simulate these real-world cases?

We thank the reviewer for their review. We agree with the reviewer that it would be useful to know which models are correct, but given the study is of a hypothetical SAI deployment and so there are no real-world cases to directly compare to then this is a very complex question. We have done a careful assessment of the model responses including an evaluation of transport against observations as well as a consideration of particular aspects of model schemes (e.g. degree of implementation of heterogenous halogen chemistry on sulfate) to help answering this complex question. In part 1 (Visioni et al., 2022) of this study, we have listed some of the evaluations against past volcanic eruptions performed with versions of the models used here (i.e. Mills et al., 2017; Clyne et al., 2021; Dhomse et al., 2021; Quaglia et al., 2022). However, what these studies often highlight is that the precise meteorological conditions at the time of the eruption strongly influence the sulfate plume evolution, and that the complex interaction of multiple factors prevents a precise assessment of the differences between models in one single aspect (such as large scale stratospheric transport). A continuous injection setup, as in our study, helps highlighting models' differences in a way that is similar, but also complementary, to an evaluation of volcanic analogues.

We have now expanded Section 4 that discusses the issue raised by the reviewer explicitly: "This thus suggests that certain degree of caution is needed in interpreting the results of studies conducted with single models, and that more work should be undertaken to improve the models and evaluate them against the available observational data, e.g. from recent volcanic eruptions to evaluate the model aerosol microphysics or using long-lived tracers to evaluate model transport. For modelling intercomparisons, understanding and attributing the reasons behind the inter-model spread rather than focusing only on the multi-model mean responses would help identify which model responses are likely more trustworthy and representative of the uncertainty in a hypothetical real-world SAI response, and which arise from spurious model features or problems with the code. This in turn would help to identify the areas in need of potential future model development and, thus, to narrow the uncertainties in future model projections of SAI impacts."

In any case, the authors need to address the points below and the 54 comments in the attached annotated manuscript.

We thank the reviewer for helpful suggestions for improving the manuscript. We address individual comments below in blue; the 54 comments in the attached annotated manuscript have also been addressed.

In addition to Fig. 10, provide ones for separate winter and summer seasons, so we can see how the polar vortex behaves in different seasons. For Fig. 10, why does 15°S injections give the largest change in the Northern Hemisphere (NH) polar vortex?

We agree with the reviewer that seasonal changes in polar vortices would be an interesting aspect to consider but believe that the current length of simulations (10 years) means that at present this is unlikely to show reliable signals due to the strong contribution of natural interannual variability. Thus in order not to confuse the reader we prefer to stick to showing the yearly mean changes only. Even for yearly mean changes, we believe the length of simulations is too short to confidently assess the dependence of the stratospheric NH polar vortex response on the latitude of injection.

We have however now expanded Section 3.5 to include: "In the extra-tropical stratosphere, CESM2, UKESM and GISS-MATRIX all simulate strengthening of stratospheric jets in both hemispheres, consistent with geostrophic balance and the strengthening of the horizontal temperature gradient brought about from heating in the lower stratosphere. The results suggest impacts on the modes of high latitude variability, including the Northern and Southern Annular Modes (NAM and SAM, respectively), which would influence regional mid- and high latitude surface temperature and precipitation responses during dynamically active seasons (e.g. winter in the NH). However, here the derived responses are substantially affected by interannual variability due to the short length of the integrations; this prevents confident analysis of any inter-model differences or the dependence of the stratospheric polar vortex response on the latitude of injection."

We have also expanded the last sentence of the Section 4 to include more stress on the need to explore the dynamical response, both in the stratosphere and in the troposphere, and its dependence on the latitude of injection in future follow up studies utilizing longer simulations: "However, given the short length of the simulations detailed analysis of the dynamical response, both in the stratosphere and in the troposphere, and its dependence on the latitude of SAI alongside the underlying mechanisms is beyond the scope of this study, but will be explored in the future with longer simulations."

In Fig. 8, why is south polar ozone depletion for 15°N and 30°N injections even larger than for SH injections?

We do not think this is true – Antarctic ozone depletion in CESM2 and GISS-OMA is largest for the injections in SH. As discussed in the manuscript, UKESM includes only a very limited representation of heterogenous halogen chemistry on sulfate and, thus, does not provide a reliable estimate of SAI impacts on the Antarctic ozone. GISS-MATRIX, on the other hand, shows Antarctic lower stratospheric ozone reduction and changes in ClO that are similar in magnitude across the different injection locations (despite the large differences in sulfate distribution) which also suggests problems in its chemistry scheme and thus does not provide a reliable estimate of Antarctic ozone depletion.

For all the figures with rows for different latitudes of injection, it would be much more intuitive if the rows with the most northerly injections were at the top of the page, that is put 30°N first, then 15°N, and so on.

We agree with the reviewer and have changed the plots.

For Figures 1, 5, 6, 7, 9, 10, and S4-S7 what is the significance of the plots. Since these are means of 8 years and 3 ensemble members, plot dotted shading over the insignificant parts.

We agree with the reviewer and have added statistical significance to the plots

For Figures 1, 5, 6, 7, 9, 10, and S4-S7 use one large color bar for entire figure and delete all the tiny illegible ones beneath each panel.

We agree with the reviewer and have changed the plots.

The color scale for Figs. 6, 7, 9, 10, and S4-S7 for negative values is ugly and counterintuitive. Use just gradually darker blue and then purple for more negative values. Don't use green.

We have changed the plots

Fig. S7 has a caption that does not agree with the figure. It looks like a time series and not an 8-year mean. If so, the x-axis needs to be labeled correctly, in time, with yearly indications and not just arbitrary numbers of months. Also, the image is very blurry and needs to be replaced with a clear one. Also, there are too many black contours that completely cover the shading and information. Use a larger contour interval. And use a better color scale.

We have now changed the figure caption.

For supplemental information, add a table of contents on the first page with a list of the tables.

Added now.

REVIEW #2

This manuscript details differences between simulations of stratospheric aerosol injection (SAI) among three models, here focusing on SAI responses in layers of the atmosphere well above the surface, with a large focus on circulation and chemistry. The in-depth exploration of model differences and their causes herein is commendable, and is very on topic for the special issue. Attempts to evaluate geoengineering strategies with simulations have been plagued by poorly understood disagreements between models. Hence, this assessment could be useful for future attempts to quantify SAI uncertainties, as well as to inform model developers and users intending to explore SAI and related scenarios. However, the manuscript explores the different responses among models without much attempt to communicate why these responses matter and why the intermodel disagreements are worth assessing. The manuscript thus in its current form presents itself more like an academic exercise than the scientific contribution it clearly could and should be. The reviewer requests major revisions, most critically textual changes to clarify the context and significance of findings.

We thank the reviewer for their review as well as helpful suggestions for improving the manuscript. We address individual comments below in blue.

General comments:

The text should be augmented to explain the significance of its findings. There's no mention of the how the stratospheric and free-tropospheric responses detailed herein could matter to humans, ecosystems, to what extent SAI is a viable strategy, or to how SAI might be designed to minimize risks and uncertainties (given the focus on varied-injection-latitude SAI experiments). Contrary to the final line of the abstract (line 48), this study does not really explore "climate impacts from SAI" in the typical sense – temperature and precipitation changes at Earth's surface, which is instead discussed only in PART1. The significance of this study, PART2, would be far clearer if the text links the stratospheric and free-tropospheric changes it analyzes to their possible ramifications for surface climate, citing relevant figures in PART1 as it does so. As is the current study only makes two passing references to the surface temperatures in PART1 (lines 224 & 230), leaving the reader with an uncommonly large amount of detective work to appreciate the significance of the results. This study should be amended to spell out the significance of the stratospheric and free-tropospheric responses it focuses on, including a handful of comments on how the circulation and chemistry features discussed here might impact the surface climate responses presented in PART1.

We thank the reviewer for these useful suggestions – we agree and have modified the manuscript accordingly, adding more context as to why changes in stratospheric and free tropospheric composition (ozone, water vapour), stratospheric temperatures and circulation are relevant for the surface climate response to SAI, adding references to PART1 when appropriate. These changes are also detailed in the specific comments and responses below.

Related to the first point, the study should clarify the links between model results and real world implications, as this is central to its purpose. Line 64-5: "[M]odel intercomparisons are useful in understanding uncertainties in climate responses to SAI". There is truth in this statement but it should be explained to the reader, as it is central to the significance of this study yet is not trivial. Under the assumptions that models differ because of poorly constrained parameters / process rates and the collection of models sufficiently samples these uncertain inputs, the spread of results across models can be used as a proxy for uncertainties in the real world (here being uncertain outcomes of SAI). But if intermodel differences are instead caused by bugs or identifiable biases, their use as a proxy for real world uncertainties is diminished – nevertheless, identification (and ultimately correction) of these issues is an important step toward the original purpose (hence this work has a technical purpose that supports the main scientific one). Running experiments in three/four models is a commendable effort though too few to truly cover the uncertainties, while the presence of model bugs discovered here complicates the applicability of simulation results to the real world. Fortunately, as is demonstrated, the number of models is sufficient to identify major differences and facilitate comments on their causes, giving this work value. Some discussion along these lines should be added as context for readers who are not climate modelers, to help them understand the results of this study.

We agree and have added a discussion of this into the manuscript, see the response to the specific comments below.

The paper should indicate which model results are more and less reliable estimates of SAI responses that would occur in the real world. Out of the four models used, some are clearly less useful than others for some purposes. Most glaring is the use of the GISS OMA scheme, which should be explained upfront. Due to the lack of interactive aerosol size this model cannot be expected to be as realistic as the others for SAI experiments, wherein aerosol size is expected to grow to far greater size than as emitted. As is a reader might naively treat it initially as on equal footing with the other, more fully interactive aerosol models. Perhaps GISS OMA's use here is meant to represent early models used in interactive aerosol geoengineering experiments that similarly had fixed aerosol size? Or more generally as a benchmark to demonstrate the necessity of the more interactive two-moment schemes? The reason for its use should be presented upfront, as its divergence from other model results is more of an expectation than a novel finding as presented here. Similarly, if "GISSmodal" (which is an inappropriate name for this model, as explained below) has a large ozone bias, the ozone biases from CESM2 and UKESM should be treated as the best approximation of the real world ozone response. The manuscript would be best if it made statements based on the collection of models deemed trustworthy for each response, rather than predominantly explaining separately what the response is in model 1, then model 2, then model 3.

As suggested, we now expand on why do we include GISS-OMA simulations in Section 2.1: "The inclusion of simpler GISS-OMA simulations in addition to GISS-MATRIX can be used as a benchmark that allows us to test the importance of detailed representation of aerosol processes for the simulated response. It is also more representative of models used in early geoengineering studies (e.g. Robock et al., 2008; Pitari et al., 2014)." "

In terms of specifying which model results are more less reliable estimates of real-world SAI responses, we believe that the current manuscript already clearly states which model results arise from problems with the models (e.g. most of GISS-OMA results or the Antarctic ozone response in UKESM) and, thus, are unlikely to be good estimates of real world SAI responses. For other instances, we tried to attribute model responses to particular model characteristics, e.g. climatological transport or size distribution of aerosols produced. Given the lack of real-world SAI examples/data to compare to, more confident discussion of which model responses are likely most reliable estimates of real-world SAI responses is very difficult and as such beyond the scope of this manuscript.

GISS ModelE with MATRIX is not a “modal” scheme despite having similarities, so “GISSmodal” should be replaced with “GISSmatrix” (in both PART 1 and 2, for consistency). As explained in Bauer et al 2008 (https://doi.org/10.5194/acp-8-6003-2008), MATRIX is based on the Quadrature Method of Moments, which “provides a computationally efficient statistically-based alternative to modal and sectional methods for aerosol simulation that does not make a priori assumptions about the shape of the size distribution”. What causes confusion is that outside the core equations of this scheme, MATRIX then treats each of its aerosol populations as lognormal size distributions, making its output look a lot like that from a modal method (but with a larger number of aerosol populations). For PART1 and PART2 to refer to “GISSmodal” would be quite bad in that this would make the distinction even more confusing for anyone who comes across these studies. Similarly, text in the manuscript that discusses “modal models” with MATRIX in this category should rename the category to “two-moment models”, as what these models share is more accurately that they enable both aerosol mass and size to change (two-moment), as compared to mass only (one-moment, as with OMA). “GISSmatrix” is the best renaming option, as it clearly connects to other uses of the MATRIX scheme in the literature, eg for AEROCOM and CMIP. Another potential renaming option might have been “GISStwo-moment”, but this would create confusion with the separate TOMAS two-moment scheme option in GISS ModelE. While not as incorrect as “GISSmodal”, “GISSbulk” would be better replaced by “GISSoma” in order to also connect more clearly to other literature using the OMA scheme.

We agree and have changed GISSmodal to ‘GISS-MATRIX’, and GISSbulk to ‘GISS-OMA’. We have also changed references to CESM, UKESM and GISS-MATRIX all having ‘modal schemes’ to ‘two-moment schemes’.

The relevance of the off-equatorial experiment setup should be explained more, as it currently seems to be an afterthought despite its central placement in the title. One would anticipate that, given the title, the results of this study are meant to guide SAI injection strategy, but there’s no statement of what the results imply for what injection latitudes are less problematic or have less uncertainties than others. Please make a statement on this. Presumably such a statement would take into the consideration the findings of both PART1 and PART2.

We agree and have now included the following paragraph as a final paragraph of the manuscript: “Finally, our results further confirm the need to think of potential SAI deployment considering multiple injection locations outside of the equator. Injecting SO2 at the equator gives rise to the lowest efficiency of global cooling per AOD (PART1) as the result of the confinement of sulfate inside the tropical pipe (thereby reducing the AOD global coverage; PART1 and Section 3.1 here) as well as leading to the largest increases in lower stratospheric temperatures (Section 3.2). The latter lead to the strongest increases in tropical lower stratospheric water vapour (Section 3.4) and ozone (Section 3.3) which act to partially offset the direct aerosol-induced surface cooling as well as can cause the strongest perturbations of stratospheric and tropospheric circulation (Section 3.3 and 3.5), thereby indirectly affecting the surface temperature and precipitation responses discussed in PART1.”

As written the abstract is more of a summary than an abstract and includes substantially more technical details than necessary, obscuring its usefulness. It likewise gives almost no indication why these findings matter. The abstract should be rewritten to clearly emphasize what the main intentions of the study are within its first few lines, and should more succinctly summarize the scientific results and their merit (eg key areas where models agree vs disagree and why these matter for potential plans to use SAI).

We agree and have now re-written the abstract to read:

"The paper constitutes part 2 of a study performing a first systematic inter-model comparison of the atmospheric responses to stratospheric aerosol injection (SAI) at various single latitudes in the tropics, as simulated by three state-of-the-art Earth System Models - CESM2(WACCM6), UKESM1.0, and GISS-E2.1-G. Building on part 1 (PART1, Visioni et al., 2022) we demonstrate the role of biases in the climatological circulation and specific aspects of the model microphysics in driving the inter-model differences in the simulated sulfate distributions. We then characterise the simulated changes in stratospheric and free-tropospheric temperatures, ozone, water vapour and the large-scale circulation, elucidating the role of the above aspects to the surface SAI responses discussed in PART1.

We show that the differences in the aerosol spatial distribution can be explained by the significantly faster shallow branches of the Brewer Dobson circulation in CESM2, a relatively isolated tropical pipe and older tropical age-of-air in UKESM, and smaller aerosol sizes and relatively stronger horizontal mixing (thus very young stratospheric age-of-air) in two GISS versions used. We also find a large spread in the magnitudes of the tropical lower stratospheric warming amongst the models, driven by microphysical, chemical and dynamical differences. These lead to large differences in stratospheric water vapour responses, with significant increases in stratospheric water vapour under SAI in CESM2 and GISS that were largely not reproduced in UKESM. For ozone, a good agreement was found in the tropical stratosphere amongst the models with more complex microphysics, with lower stratospheric ozone changes consistent with the SAI-induced modulation of the large-scale circulation and the resulting changes in transport. In contrast, we find a large inter-model spread in the Antarctic ozone responses that can largely be explained by the differences in the simulated latitudinal distributions of aerosols as well as the degree of implementation of heterogeneous halogen chemistry on sulfate in the models.

The use of GISS runs with bulk microphysics demonstrates the importance of more detailed treatment of aerosol processes, with contrastingly different stratospheric SAI responses to the models using the two-moment aerosol treatment; however some problems in halogen chemistry in GISS are also identified that require further attention. Overall, our results contribute to an increased understanding of the underlying physical mechanisms as well as identifying and narrowing the uncertainty in model projections of climate impacts from SAI."

Specific comments:

Lines 21, 91, 119, 333: The injection latitudes are unnecessarily listed four times. Please remove at least two of these instances to be less repetitive.

As suggested, we have removed the first two occurrences of this (i.e. in the abstract and at the end of Section 1).

Lines 46-8: “[O]ur results contribute to an increased understanding of [...] the sources of uncertainty in model projections of climate impacts from SAI”. This is a critical statement of purpose yet as written encloses itself in ‘model land’, neglecting the greater goal to further understanding of SAI uncertainties as they would matter in the real world. Please rework this to convey this study’s significance in understanding how the real world might be under the hypothetical SAI scenarios.

As noted above, we have now re-written the abstract.

Lines 55-6: Please briefly state why these “side-effects” are important and worth study.

We have now added “These side-effects can thus modify the direct response to SAI, further modulating the radiative balance as well as impacting regional climate and ecosystems.”

Lines 64-5: Yes, “models are themselves imperfect”, but this does not in itself make them useful for understanding uncertainties, as stated (the expectation would be that model imperfections make them not useful). Please rework this in consideration of the general comments (second paragraph).

As suggested we have now added the following discussion to the text: “Such uncertainties arise from many sources, including the efficiency of SO2 to aerosol conversion, the extent to which sulfate aerosols will be transported away from the injection locations by the large-scale circulation and mixing processes and the removal of the aerosols from the atmosphere altogether, the efficiency of the direct impacts of aerosols on the radiative balance as well as from uncertainties in any indirect impacts, for instance on atmospheric circulations and clouds. Simulations with a number of different models can thus help represent the uncertainty in real world climate response to a hypothetical SAI, whilst identifying and attributing certain characteristics of individual model responses to particular aspects of model design or features can help in narrowing this real world uncertainty.”

Line 107: Three points here. First, “the key findings”, not “they key findings”. Second, what “key findings”? Third, the introduction would be stronger if its last sentence were a more general segue to the next section.

As suggested, we have added an outline of the rest of the manuscript to the end of Section 1: “Section 2 summarises the model simulations performed. In Section 3.1 we focus on the simulated sulfate aerosol distributions, and evaluate and discuss the role of biases in model transport in contributing to the inter-model spread. We then discuss the associated SAI impacts on stratospheric temperatures (Section 3.2), ozone and the large scale residual circulation (Section 3.3), water vapour (Section 3.4) and zonal winds (Section 3.5). Finally, Section 4 summarises and discusses the main results. ”. We also changed ‘the key findings’ to ‘the main findings’.

Lines 110-123: The methods section is awkwardly quite short for the obvious reason that nearly everything is in PART1. Perhaps some of the material in the “Results” section actually belongs here and should be moved? I’m thinking particularly of the SAD diagnostics (Lines 130-139).

As suggested, we have now moved the description of the SAD diagnostic to Section 2.1.

Line 115: Please explain here why GISS OMA is used. Its lack of dynamic aerosol size would seem inappropriate for a study where so much sulfur is emitted that the aerosol would be nowhere like their fixed emission size (see discussion of this in general comments, paragraph three).

The text now says: “The inclusion of simpler GISS-OMA simulations in addition to GISS-MATRIX can be used as a benchmark that allows us to test the importance of detailed representation of aerosol processes for the simulated response. It is also more representative of models used in early geoengineering studies (e.g. Robock et al., 2008; Pitari et al., 2014).”

Line 115: GISS ModelE’s MATRIX is not a “modal” scheme so the name and description of “GISSmodal” should be altered to clarify this (see general comments, paragraph four).

We agree and have changed GISSmodal to ‘GISS-MATRIX’, and GISSbulk to ‘GISS-OMA’

Lines 115-6: Please alter or add to this sentence to instead trace out how the methods (models + custom output) are used to achieve this manuscript’s main aims (see general comments, first three paragraphs). To “test the importance of detailed representation of aerosol processes for the simulated response” doesn’t seem like a major goal, unless one of the main goals is to argue the 3.

That line referred to the use of GISS-OMA in addition to GISS-MATRIX simulations, and not to all of the objectives of the study. We have now expanded this part to avoid future confusion and include both: “The use of three ESMs allows us to better constrain the uncertainty in the climate response to SAI. The inclusion of simpler GISS-OMA simulations in addition to GISS-MATRIX can be used as a benchmark that allows us to test the importance of detailed representation of aerosol processes for the simulated response. It is also more representative of models used in early geoengineering Robock et al., 2008; Pitari et al., 2014

Lines 137-8: Where are the “2” in Eqn. 1 and the “4.5” in Eqn. 2 from? Lognormal statistics? Please specify.

They directly derive from the calculation of the i-th raw moment of a lognormal distribution. For m_2 (SAD) the factor ends up being 2, for m_3, 9/2

Line 138. The “r” in Eqn. 2 looks like it should instead be an “ri”.

The reviewer is right – we have now corrected this.

Lines 176, 193, 215, 217, 235, 242, 296, 310, 314, 358, 370, 377, 389, 422: All these lines generalize the 3 models other than GISS OMA as “modal”, despite “GISSmodal” being a misnomer for what is not properly a modal model. More appropriately the commonality is that these are “two-moment” models, representing both mass and number as changing rather than just mass.

We agree and have changed these to the two-moment aerosol schemes.

Line 188: Cite that the QBO response is focused on later in the manuscript (within Section 5.3.5).

Done.

Line 202: It would be surprising if the LW effect is indeed substantially aerosol-size sensitive, as is more established with the SW effect. Does the Laakso et al 2022 really show this? Their Section 3.1.1 raises other reasons (differences in optical properties and radiative transfer schemes).

Laasko et al. indeed shows this (see for instance Fig. 2 in that manuscript, magenta line. The 8000 nm wavelength absorption grows substantially right at the sizes where microphysical growth matters in SAI studies.

Lines 235, 272, and 301: The sectional breakdown between 3.3.1, 3.3.2, and 3.3.3 seems clunky. 3.3.1 sounds like it contains nearly all the info ("tropics" and "mid-latitudes"?). Perhaps 3.3.1 and 3.3.2 should be combined into a single "Stratospheric ozone" section in contrast to what's now 3.3.3 ("Tropospheric ozone changes").

We have renamed Section 3.3.1 as "Stratospheric ozone changes in models with two-moment microphysics".

Line 237: Please briefly explain why stratospheric ozone response to SAI is important. Is it that this matters to human health via impacts on UV radiation? Might ozone's role as a greenhouse gas impact the surface cooling effectiveness as shown in PART1?

The reviewer is correct – we have added this info to the manuscript: "The absorption of incoming solar radiation by stratospheric ozone plays a crucial role in shielding the Earth surface from the harmful UV radiation, thus having direct impacts on human health and ecosystems. In addition, the absorption of outgoing terrestrial radiation by ozone in the troposphere and lower stratosphere contributes to the greenhouse effect. Therefore, any ozone changes there can modulate the direct radiative response from aerosol reflection, impacting the surface temperature responses discussed in PART1."

Line 258: Missing a period.

Corrected.

Line 272: Please briefly state why one should care about the SAI ozone response specifically in the Antarctic stratosphere.

We have added: "Previous decades have seen significant reductions of ozone in the Southern Hemisphere (SH) high latitudes brought about by accelerated heterogenous halogen reactions inside the Antarctic polar vortex as the result of anthropogenic emissions of ozone depleting substances. And so future evolution and recovery of Antarctic ozone continues to be the focus of significant scientific and political interest (WMO, 2018),"

Line 301: Please clarify why tropospheric ozone response to SAI is important, and why it should be discussed separately from stratospheric ozone.

We have added: "In addition to acting as a greenhouse gas, in the troposphere ozone constitutes an atmospheric pollutant, adversely impacting human health (e.g. Eastham et al., 2018), crop production (e.g. Xia et al. 2017) and ecosystems (e,g. Zarnetske et al., 2021)."

Line 322: Why does stratospheric water vapor response to SAI matter? Is this through chemistry influence that itself matters for health via ozone/UV or feedbacks on surface climate? Or water vapor acting as a greenhouse gas in a way that itself alters the radiative forcing, and hence SAI efficiency? This should briefly be explained.

"As with ozone, the absorption of outgoing terrestrial radiation by water vapour in the lower stratosphere and the troposphere contributes to the greenhouse effect. Thus, any SAI-induced changes in it can further modulate the radiative balance and surface temperature responses discussed in PART1. In addition, the photolysis of stratospheric water vapour (SWV) constates the main source of reactive HOx in the stratosphere, which act to reduce stratospheric ozone levels and thereby further modulate the ozone responses discussed in Section 3.3."

Line 342: This section (3.5) details features of "zonal winds" under SAI, yet makes no reference to the Northern and Southern Annular Modes, the most frequently discussed zonal wind structures in the stratosphere. Changes to these structures are apparent in Fig. 10, so may deserve explanation in the section. If the simulations are too short for interpretation to be worthwhile (mentioned vaguely in lines 442-4), explain here this decision rather than ignoring NAM and SAM entirely.

We have now expanded this paragraph to read: "In the extra-tropical stratosphere, CESM2, UKESM and GISS-MATRIX all simulate strengthening of stratospheric jets in both hemispheres, consistent with geostrophic balance and the strengthening of the horizontal temperature gradient brought about from heating in the lower stratosphere. The results suggest impacts on the modes of high latitude variability, including the Northern and Southern Annular Modes (NAM and SAM, respectively), which would influence regional mid- and high latitude surface temperature and precipitation responses during dynamically active seasons (e.g. winter in the NH). However, here the derived responses are substantially affected by interannual variability due to the short length of the integrations; this prevents confident analysis of any inter-model differences or the dependence of the stratospheric polar vortex response on the latitude of injection."

Line 356: Replace "a lot of" with less casual wording (eg "extensive").

We now say "derived responses are substantially affected by interannual variability"

Line 356-7: "due to the short length", not "due to short length".

Corrected.

Line 357: "prevents", not "prevent"

Corrected.

Lines 409-44: It would be well worth going over these paragraphs to ensure they advertise the manuscript's best scientific and technical strengths, since there's room for improvement here (see general comments).

We have now combined and reworked two of these paragraphs to read:

"Our findings illustrate the importance of a detailed and adequate representation of a range of microphysical, dynamical and chemical processes in models for accurately representing the potential impacts from SAI, both directly in the stratosphere as well as lower down at the surface. By demonstrating the role of biases in climatological circulation, our results highlight the importance of not only model microphysics but also transport processes for simulating the evolution of the aerosol plume. They also highlight the large uncertainties in the representation of these processes in current Earth System Models and the need for realistic representation of both aspects for determining the aerosol response and, thus, the potential impacts of SAI on atmospheric radiative balance, composition and circulation. This thus suggests that certain degree of caution is needed in interpreting the results of studies conducted with single models, and that more work should be undertaken to improve the models and evaluate them against the available observational data, e.g. from recent volcanic eruptions to evaluate the model aerosol microphysics or using long-lived tracers to evaluate model transport.. For modelling intercomparisons, understanding and attributing the reasons behind the inter-model spread rather than focusing only on the multi-model mean responses would help identify which model responses are likely more trustworthy and representative of the uncertainty in a hypothetical real-world SAI response, and which arise from spurious model features or problems with the code. This in turn would help to identify the areas in need of potential future model development and, thus, to narrow the uncertainties in future model projections of SAI impacts. ”

Line 409-11: Is this statement an accurate representation of the study? Given this study is wholly assessing models of hypothetical SAI scenarios (no observations), it doesn't thoroughly comment on what is "realistic". As expressed in the general comments, the goals of this study should be presented more clearly.

We have changed this to "the importance of a detailed and adequate representation of a range of microphysical, dynamical and chemical processes".

Lines 434-6: This is a restatement of lines 46-8, so please see the comment for those lines. This would be an appropriate place for discussion of just how applicable the intermodel spread is to understanding SAI uncertainties, as they would relate to actual deployment in the real world (see general comments, paragraph 2).

We have now reworked this paragraph to read "For modelling intercomparisons, understanding and attributing the reasons behind the inter-model spread rather than focusing only on multi-model mean responses would help identify which model responses are likely more trustworthy and representative of the uncertainty in a hypothetical real-world SAI response, and which arise from spurious model features or problems with the code. This in turn would help to identify the areas in need of potential future model development and, thus, to narrow the uncertainties in future model projections of SAI impacts."

Line 439: Does this locking of the QBO matter? Should this inform what latitudes SAI should be injected at? Please make a concluding statement on (or at least discussion of) which latitudes should or should not be used, considering the relative importance of this QBO locking to other injection latitude sensitivities from both PART1 and PART2. This would seem to be the natural wrap up for the "off-equatorial" injection focus in the study's title.

Regarding the role of QBO and impacts of QBO locking, we have now included:

"In general, the variability in equatorial zonal winds has been linked to variability in tropical tropospheric convection, subtropical and mid-latitude tropospheric jets as well as modes of high latitude variability, e.g. North Atlantic Oscillation (Anstey et al., 2022). Therefore, SAI impacts on the QBO, including locking in a permanent westerly phase under equatorial injections, have a potential to impact the circulation in regions outside the equatorial stratosphere, although longer simulations would be needed to confidently diagnose such teleconnections."

As noted in the response to the main comments above, we now include a discussion of the relevance of the off-equatorial set-up as a final paragraph of Section 4: "Finally, our results further confirm the need to think of potential SAI deployment considering multiple injection locations outside of the equator. Injecting SO_2 at the equator gives rise to the lowest efficiency of global cooling per AOD (PART1) as the result of the confinement of sulfate inside the tropical pipe (thereby reducing the AOD global coverage; PART1 and Section 3.1 here) as well as leading to the largest increases in lower stratospheric temperatures (Section 3.2). The latter lead to the strongest increases in tropical lower stratospheric water vapour (Section 3.4) and ozone (Section 3.3) which act to partially offset the direct aerosol-induced surface cooling as well as can cause the strongest perturbations of stratospheric and tropospheric circulation (Section 3.3 and 3.5), thereby indirectly affecting the surface temperature and precipitation responses discussed in PART1."

Line 440: Missing comma, "[…] of the simulations, detailed […]" .

Corrected

Lines 442-4: Please clarify what dynamical responses are being skipped due to the short simulation length. Please also state the rough number of years or decades that would be needed.

We have now clarified this: " However, given the short length of the simulations, detailed analysis of the dynamical response, both in the stratosphere and in the troposphere (e.g., impacts on the Northern and Southern Annular Modes), and its dependence on the latitude of SAI alongside the underlying mechanisms is beyond the scope of this study, but will be explored in the future with longer simulations (and multiple ensemble members)."

Figs. 5-7,9,10: Please revise each figure to have only one large colorbar to avoid overcrowding.

We agree and have changed the plots.

Fig. 6: What does the black contour shading represent? This is not on the colorbar.

As explained in the figure caption, "Contours show the vertical velocities in the control SSP2-4.5 run for reference." We have now adjusted the contours in the panels showing GISS responses to make them clearer, i.e. we only plot the zero line contour.

Fig. 8: Please include only one legend on the figure, which should be larger.

We agree and have changed the plots.