Response to reviewer 2

Summary: This paper presents the first multi-model comparison of Earth-system model simulations of high-latitude low-altitude (HiLLA) stratospheric aerosol injection (SAI) using CESM2-WACCM, UKESM1, and E3SMv3. HiLLA SAI is proposed as a potential method of climate intervention which could serve as an early-stage deployment option since existing aircraft could be repurposed for deployment, in contrast to SAI deployment methods which consider lower latitude and higher altitude injection and require the development of new aircraft. The results of this manuscript assess the global and regional responses of temperature, precipitation, aerosol optical depth, sulfate burden, sea ice, sulfur deposition, and stratospheric ozone to idealized 35-year HiLLA SAI deployment simulations which inject 12 Tg SO2 per year into the polar stratosphere, with injections occurring at 60oN from March through June, and at 60oS from September to December. Sensitivity to injection height (13 vs. 15 km in altitude), altitude, seasonality, and longitude are considered. The main finding of this work is that HiLLA SAI deployment could reduce global temperatures. Additionally, HiLLA SAI deployment is found to induce strong polar cooling, increased sea-ice area, decreased high-latitude precipitation, and an increased Arctic seasonal cycle, relative to the background climate change scenario and to a low-latitude high-altitude SAI deployment scenario. However, the cooling efficiency, defined as the magnitude of near-surface air temperature change per unit of injected SO2 , is reduced in HiLLA-SAI scenarios compared to traditional high altitude injection strategies. The simulations and key results presented in this study contribute to a growing body of literature evaluating possible SAI deployment strategies. The analysis of the global and regional climate impacts of the HiLLA SAI simulations, as well as the sensitivity testing of HiLLA SAI to injection seasonality, latitude, longitude, and altitude are overall comprehensive and well done. However, there are a few major comments and several minor comments that should be addressed prior to publication.

We thank the reviewer for their positive, helpful and thorough comments, which will substantially improve our paper. See below for point-by-point responses.

Major Comments:

The authors should consider adding in some more commentary on some of the major differences between the considered models (CESM2-WACCM, UKESM1, E3SMv3), particularly when relevant to inter-model differences in the climate response to HiLLA deployment. - One example when this is already done nicely is in Section 3.7, with the discussion of changes in stratospheric ozone, CESM2-WACCM is the only model that experiences decreases in stratospheric ozone in response to HiLLA SAI deployment since this model includes heterogeneous chemistry in the stratosphere, whereas UKESM1 and E3SMv3 do not.

Thanks for this useful suggestion. We will add more discussion of structural differences between the models which contribute to variation in outcome where possible in the revisions described below. First, in the discussion of Fig 4, we will describe the differences in SSP2-4.5 zonal mean warming patterns (related to variation in the strength of Arctic amplification and AMOC decline) in UKESM1 and CESM2-WACCM, which lead to the differences in Fig 4k. We will add the following text after line 193:

> *“The large differences between UKESM1 and CESM2-WACCM in the zonal temperature changes relative to the nominal `target' state arise, at least in part, from differences in the meridional pattern of warming in these two models. UKESM1 has strong Arctic amplification of warming, which is not fully offset with injection at 30°N under the ARISE scenario (Henry et al., 2023). Whereas CESM2-WACCM has less strong background Northern Hemisphere high latitude warming, in part due to a strong North Atlantic warming hole (and associated decline in the Atlantic Meridional Overturning Circulation), which is reduced in UKESM1 (Henry et al., 2023).”*

- Lines 223-233 discuss AOD per unit injection and global mean cooling per unit global mean AOD. Why might UKESM1 differ from CESM2-WACCM and E3SMv3 in AOD per 10 Tg injection and change in global mean temperature per 0.1 AOD?

Variation in AOD per unit injection arises from a combination of variation in aerosol lifetime (and therefore burden of SO4 per unit injection) and of variation in size distribution (and therefore AOD per unit burden). Here, the higher UKESM1 AOD per unit injection must be driven by aerosol size, since UKESM1 also has lower burden per unit injection than the other models. In UKESM1, the sulfate aerosols are significantly smaller than in CESM2-WACCM (Fig A6), and UKESM1 has smaller aerosols than other models in other recent SAI simulations (Lee et al., 2025). The cause of the lower global mean cooling per unit AOD in UKESM1 than the other two models, which is not seen (relative to CESM2-WACCM) in the ARISE simulations, is not clear, but may be related to the weak Antarctic cooling under HiLLA scenarios in UKESM1.

We will add a discussion of this at line 228 as follows:

Variation in AOD per unit injection arises from a combination of variation in aerosol lifetime (and therefore burden) and of variation in size distribution (and therefore AOD per unit burden). Here, the higher AOD per unit injection in UKESM1 is likely driven by UKESM1’s smaller aerosol size (Figure A6), since UKESM1 also has lower SO4 burden per unit injection than the other models (Figures 7 and 8). The cause of the lower global mean cooling per unit AOD in UKESM1 than the other two models (which is not seen in the ARISE simulations) is not clear, but may be related to the weak Antarctic cooling under HiLLA scenarios in UKESM1 (Figure 2).

- Lines 250-260 assess the impact of varying injection latitude, longitude, and seasonality. Why might there be so much inconsistency between considered models, particularly when considering the latitude of injection?

The causes of these inter-model differences are not fully explained by our analysis, but are potentially driven by differences in the background circulation in the models. We will add a note on this to the discussion (at line 391) as follows:

“Future work should explore the drivers of this inter-model difference in the response to variation of injection latitude, particularly as it relates to variation in the models' background lower stratospheric circulation.”

The discussion of Figures 7 and 8 should be considered carefully. Figure 7 shows only results from UKESM1 and CESM2-WACCM. Figure 8 shows only results from UKESM1 and E3SMv3. This should probably be noted in the text. Why might the majority of the sulfate be located equatorward of the injection location?
We will add a note explaining the data limitations which mean we only include 2 out of 3 models in figure 7 as follows: (at line 220) "Due to differences in data outputs between the models, Figure 7 shows results only for UKESM1 and CESM2-WACCM".

For Figure 8, column integrated burdens are available for CESM2-WACCM, so will update this figure to use all three models.

We will also add a sentence discussing the larger equatorward burden, at line 222 as follows: *"The larger burden equatorward of the injection sites is not inconsistent with a poleward mean flow; it could be explained by a greater average lifetime of sulfate which is mixed equatorwards, given the reduced downwelling at lower latitudes."*

In section 3.6, statistical significance testing is discussed relative to a "5% false discovery rate significance threshold (Wilks, 2016)." The methods outlined in Wilks (2016) for adjusting for the False Discovery Rate could be applied to a variety of different statistical tests (e.g., student's t-test). As such, the authors should include more details on the specific statistical test where p-values were then adjusted to account for the False Discovery Rate. Additionally, the authors should elaborate on why they chose to test for significance for precipitation changes, but not for other variables like temperature and AOD, for example. The authors might consider what value significance testing brings to analysis of precipitation changes, and whether the interpretation of the analysis would be altered if it was removed, given the relatively sparse regions of significance, and considering that only a single ensemble member was conducted for each simulation. Perhaps plotting precipitation as a percent change plot relative to SSP2-4.5 or to historical data would be more informative.

Thanks for this useful point. After reconsidering, we will plot precipitation here as a % change relative to SSP2-4.5 instead of in absolute units. We included the significance testing to avoid over-interpreting the noisy signal in precipitation, which is less of an issue for temperature and AOD where the signal in the decadal mean is large relative to the variability. However, we will also add the same significance hatching to the temperature plots in Figure 4(a-f) to make this explicit. As we only have one ensemble member here, even large precipitation changes can be statistically insignificant, but we feel keeping the hatching is still more informative than removing it, given the discussion at lines 327-330. However, we will also add discussion on this point in the final section (see below). We will also add a note in the figure caption that the false discovery rate adjustment was applied to p-values from Welch's t-test (used here rather than Student's to avoid assuming equal variance between samples).

In the Discussion and Conclusions section, please add a bit more discussion on the limitations of the results presented in this study. An important example would be that there was only a single simulation run for each model. Another might be that these simulations considered HiLLA SAI in the context of one emissions scenario and only presented one idealized deployment strategy. Also in this section, consider a bit of discussion on the potential implications of injecting only in one hemisphere at a time seasonally, and what this might mean for global circulation and thus, regional climate. Given the asymmetry of this deployment strategy, this will be an important aspect of risk assessment.

We will expand the discussion section to elaborate on limitations as suggested. At line 395, we will add text as below:
"Only one ensemble member was run for each model for each HiLLA strategy here, which limits our analysis, particularly of precipitation responses for which the background variability is large. Additionally, while we assessed two main deployment strategies, as well as the sensitivity to variation in these strategies (Figure 10), this still represents only a narrow window into the total space of possible HiLLA deployments. Future work could usefully broaden this analysis, such as via consideration of dependence on the background emissions scenario, and dependence on injection species."

We will also add discussion on asymmetry in our final paragraph as follows:

"HiLLA strategies produce a more asymmetric pattern of forcing, with forcing alternating between the hemispheres in their respective summer. The consequences of this asymmetry are briefly explored in Figures A2 and A3. These show that despite summer-hemisphere-only forcing, our simulations do not show a suppression of the seasonal cycle of zonal mean temperature (in most latitudes, Figure A2) or the interhemispheric temperature gradient (Figure A3), relative to the SSP2-4.5 background. Future work should assess the broader circulation impacts of summer-hemisphere-only forcing, which are not explored here."

There are some components of the figures that are not well described in the caption or in the text. Please make sure that the figure captions are completely descriptive of what is shown in each figure. Here is a non-exhaustive list of changes that should be made:

- "r1" is used to represent the first ensemble member of the SSP2-4.5 simulation in Figure 2 and Figure 3. This should be written out explicitly in the figure caption.
  We will write this out explicitly - adding the following sentence to Figure 2 caption:
  "'r1' refers to the first ensemble member of the SSP2-4.5 scenario, from which the HiLLA simulations branch."
  The reference to 'r1' in the Figure 3 caption was an error, as these differences are shown relative to the SSP2-4.5 ensemble mean, which we will now state explicitly in the caption.

- Figure 2: This is only briefly mentioned in the manuscript, and nothing about the results shown in the figure are mentioned. The authors should add a brief comment on what is shown in the figure (e.g., what is the point of having the figure in the manuscript at all?) or consider moving the figure into the supplement.

We feel that it is useful to keep fig 2 in the main body, in part to orient the reader with the simulation set-up, and will add some additional discussion of the results presented in Figure 2 from line 156 as follows:
“Given the large magnitude of injection (12 Mt in total) from year one of deployment, decreases in Global and Arctic temperature beyond the 5-member ensemble range of the SSP2-4.5 simulations happen within the first few years (Figure 2). In the Antarctic, the internal variability in temperature is larger and so the cooling signal takes longer to emerge. Figure 2 also demonstrates the lower efficiency and more polar cooling profile (see below) of HiLLA SAI which means that despite the large magnitude of injection used here, cooling in the Tropics equates to only a several decade delay of warming under SSP2-4.5.”

- Figure 2 caption: Mention that for UKESM1 and CESM2-WACCM, the solid gray line represents the ensemble mean under SSP2-4.5, the dashed line represents the ‘r1’ ensemble member, and that the light gray shading is the spread of the ensembles (that is what I assumed it to be at least). Additionally, for the HiLLA lines, mention that HiLLA-13 is represented by the solid colored lines and HiLLA-15 is represented by the dash-dotted lines.
We will add in these details as suggested, with the additional text as follows:
“HiLLA-13 is represented by the solid colored lines and HiLLA-15 by the dash-dotted lines. ‘r1’ refers to the first ensemble member of the SSP2-4.5 scenario, from which the HiLLA simulations branch, and is shown by the grey dashed line. The SSP2-4.5 ensemble mean (first five members) is shown by the solid grey line and the range across these members by the shaded grey region.”

- Consider splitting Figure 4 into two where one figure includes (a)-(f) and the other includes (g)-(k).
We would prefer to keep figure 4 as one figure, given the large number of figures overall in the manuscript.

- Figure A3: Even if HiLLA-13 has a “qualitatively similar pattern”, it would be informative to also include the data from this simulation in the plot.
Thanks, we will redraw this figure including lines for the HiLLA-13 scenario

- Figure A10: What is shown by the blue line in plot (b) showing the zonal mean deposition of SO4 in CESM2-WACCM? Why aren’t the zonal means of HiLLA-13 and HiLLA-15 simulations shown for this plot?
The dotted and solid blue lines here are the HiLLA-13 and HiLLA-15 zonal means, we will redraw the figure adding this to the legend.

Minor Comments:

- The authors should comb through the manuscript, figure titles, and figure captions to ensure that there is consistent labeling of models and simulations throughout. There are several different abbreviations used for the models where CESM2-WACCM is also noted as “CESM” and “CESM2”, UKESM1 is also noted as “UKESM”, and E3SMv3 is noted as “E3SM.”
Thanks for catching this - we will update all figures and mentions in the text to use ‘UKESM1’, ‘CESM2-WACCM’ and ‘E3SMv3’ throughout.

- Line 12: “For 13 km inject” → “For 13 km injection”
Thanks, we have fixed this typo

- Lines 22-25: “cooling efficiency” should be defined here, where it is first mentioned, rather than in Lines 41-42 where it is currently defined.
Thanks - we will amend this sentence to remove the reference to efficiency here, as follows: “could ~~achieve sufficient cooling efficiency to~~ be a viable climate intervention”

- Lines 43-44: Please describe briefly what is meant by “a less favorable background circulation.”
Thanks - we will add a note as follows: “*a less favourable (i.e. more downwelling) background circulation*”

- Line 55: Please indicate what some of these local feedbacks are (e.g., ice-albedo), they do not need to be comprehensively explained.
We will add some examples to this sentence as follows:
“Local feedbacks - particularly the albedo, Planck and lapse rate (Pithan & Mauritsen, 2014) - contribute to …”

- Line 62: “Beyond the overall global cooling efficiency…” → “Beyond the overall reduced global cooling efficiency…”
Agreed, we will re-write this sentence as: “*Beyond the reduced global cooling efficiency, HiLLA-SAI would result in many other important differences..*”

- Lines 87-88: What is the specific version of UKESM1 used for these simulations?
We will add a note of the specific version as follows:
“*The version of UKESM1 used here – UKESM1.0 (Sellar et al., 2019) – has been used extensively..*”

- Lines 122-123: “... ECMWF Reanalysis version 5 (ERA5) reanalysis…” → “... ECMWF Reanalysis version 5 (ERA5)...”
We will fix this as suggested.

- Line 125: Might be worth mentioning that SSP2-4.5 is a moderate emissions scenario in line with current policies and then can cite O’Neill et al. (2017).

Thanks, yes, we will note this, and cite O’Neill et al., as follows:

"All simulations are run under the background SSP2-4.5 emissions pathway (O'Neill et al., 2016), the SSP emissions scenario which most closely matches current policies (Hausfather, 2025)."

- Line 140: "we vary (independently)" → "we independently vary"

We have made this edit as suggested

- Lines 159-160: "Raising the altitude of injection to 15 km increases the global cooling efficiency by 60%, 34%, and 62% in UKESM1, CESM2-WACCM, and E3SMv3, respectively." I would recommend changing the percent changes to the actual cooling efficiencies of the 15 km injection.

Agreed, we will change this sentence to include the absolute values as well as the percentages - as follows:

"..increases the global cooling over the final 20 years to 0.92°C, 0.93°C, and 1.04°C in UKESM1, CESM2-WACCM and E3SMv3 (these values are increases of 60%, 34% and 62% relative to the 13 km case, respectively)."

- Lines 161-162: As mentioned in the comment for Line 55, consider explicitly mentioning a few of the feedbacks which drive Arctic Amplification.

We will expand on this at line 55 (as described above), and reference that section here.

- Lines 163-164: Reference Figure 5.

Thanks, we will add a reference as suggested: *"..than global mean cooling, for the 13 km injection case (Figure 5)."*

- Lines 177-178: "We see that efficiency of cooling is comparable to or greater than ARISE over much of the high latitudes, particularly under the 15 km injection case." Rephrase this sentence. It is a bit misleading to say that cooling efficiency is "comparable or greater" than ARISE for HiLLA SAI. This is particularly so for the HiLLA-13 case. Perhaps it could be something like, "Cooling efficiency is lower compared to ARISE over most regions of the globe for HiLLA-13. For the 15 km injection case, however, there are large regions of the globe, particularly over high latitudes, where cooling efficiency is comparable or greater than ARISE."

Thanks, we will rephrase this sentence, to emphasise the fact that efficiency is lower over most regions in both cases more strongly. We will rephrase it as follows:

"The efficiency of cooling is lower than that of the ARISE scenario over most regions of the globe for both the 13 km and 15 km HiLLA cases. HiLLA-SAI is much less efficient than ARISE in the tropics and sub-tropics. With injection at 13 km, almost all regions equatorward of 40° see less than 40% of the ARISE efficiency in both UKESM1 and CESM2-WACCM. The average efficiency relative to ARISE over the tropics (23°S-23°N) is 28% in UKESM1 and 33% in CESM2-WACCM. However, HiLLA-SAI achieves efficiencies of cooling comparable to or greater than ARISE over much of the high latitudes under the 15 km injection case."

- Lines 217-218: Consider rephrasing this sentence to something like, “While the largest burden of sulfate occurs at the poles, there is also transport equatorward and upward.” Take care of use of the word “significant.”
Thanks, agreed, we will rephrase this as:
“While the highest concentrations of sulfate occur at the poles, there is also equatorward and upward transport”

- Line 221: “... additional atmospheric sulfate burden equatorward…” → “... additional atmospheric sulfate burden is equatorward…”
Thanks! Fixed.

- Line 236: What is meant by “fast-responding” systems?
On reconsidering, we will edit this to be more specific - it is the AOD in particular we assess here so we will state this directly:
“In addition to the two central cases, for which we present 35-year simulations, we also simulate a set of nine 3-year simulations, from which we can assess the sensitivity of AOD to the injection latitude, altitude, longitude and seasonality.”

- Line 243: Where is the change in total SO4 burden shown? Reference Figure 8 here.
The change in total SO4 burden was not shown in any figure - we will add a supplementary figure including this, and the AOD per unit SO4 burden for the HiLLA scenarios.

- Lines 281: Change “all SAI simulations” to “all previously conducted SAI simulations” or something similar
We will reword this to *“..and SAI simulations consistently show..”*

- Line 320: This sentence is a bit misleading. The regions of statistically significant changes in tropical precipitation are relatively small, particularly for CESM2 and E3SMv3. Consider removing this sentence or rewording it.
We will remove this sentence, so the paragraph will start *“One potential impact..”*

- Lines 357-361: Add a brief statement mentioning how the HiLLA-13 impacts compare to HiLLA-15.
We will add a sentence to this effect as follows:
“..during the winter months. HiLLA-13, that is, the same SAI scenario but at 13 km injection altitude, shows qualitatively similar impacts but with substantially reduced magnitude – HiLLA-13 achieves ~ 0.6°C of cooling with the same total annual injection of SO2.”

- Lines 372-373: What is meant by “this effect can be overstated”?
Thanks, we agree this phrase was ambiguous. We will alter the sentence to read as follows:
“While the HiLLA strategies produce a more polar cooling than high-altitude subtropical SAI, this effect is perhaps smaller than sometimes assumed.”

- Line 377: Explain why higher altitude of injection leads to higher cooling efficiency.

We will add a note as follows:
"..the higher altitude raises efficiency markedly by increasing the aerosol burden (Figure 7)."

- Figure 1 caption: Remove the extra "for"

Done, thanks!

- Figure 4: A (k) label is missing from panel (k) plot.

Added, thanks!

- Figure 9 caption: "10Tg" → "10 Tg"

Fixed, both in the caption and also in the titles of subplots (a) and (c). Thanks!

- Figure 12 caption: Include the definition of Arctic when it is first mentioned.

Thanks, agreed. The caption will now read *"Seasonality of Arctic (>66°N) warming and SAI cooling.."*

- Figure 13 caption: The regions and their latitude definitions should be explicitly defined in the figure caption.

We will explicitly define these regions as follows:
"Arctic" here refers to all sea ice in the Northern Hemisphere, and "Antarctic" refers to all sea ice in the Southern Hemisphere.