Reviewer comments are in **bold** and the authors' responses are in blue.

We thank the reviewer for their thoughtful and constructive evaluation of our manuscript. The central concern raised on whether CESM2-WACCM6 is an appropriate model for assessing air quality-related health impacts of SAI is an important one, and we have worked to address it directly. In the revised manuscript, we have clarified the limitations of CESM2-WACCM6 for air quality applications, drawing on previous work evaluating the model and highlighting where gaps remain. We have removed wording that overstated the comprehensiveness of the model, added explicit caveats in the abstract and discussion regarding photolysis and missing aerosol species, and reframed our conclusions. We have also expanded our interpretation of the conclusions in relation to the "climate penalty" literature, and added supplemental analysis to provide further context for some of our results. Below, we provide point-by-point responses to the reviewer's comments and describe the corresponding changes made in the revised manuscript.

Major comments

The two most significant concerns are related, and boil down to the question of whether the model being used is appropriate for the task at hand. On the one hand, the authors make a good case that the fully interactive nature of the CESM2-WACCM6 simulation means that it can capture key meteorological responses to SAI, which are likely to be significant to the air quality response. Since these responses were often either neglected entirely or crudely parameterized in previous studies, directly simulating the interactions of changing meteorology with air quality is a valuable advance. However, the first question is whether the chosen model is appropriate for simulations of air quality. I am aware of almost no studies which have used WACCM for air quality modelling, beyond one study which is cited by the authors and which was itself an intercomparison of CMIP6 models. There seem to be several modelling choices in WACCM which, while sensible for a model of whole-atmosphere climate responses, might compromise its representation of air quality responses. For example Hancock et al. (2023) indicate that WACCM does not include any representation of ammonium or nitrate aerosols, but these are standard in models such as CMAQ which are dedicated to air quality – and such aerosols can be dominant in understanding air quality responses to climate change (see e.g. Nolte et al. (2018)). Indeed it appears that a recent paper in ACP which used WACCM for the boundary conditions in an air quality simulation specifically chose to use WRF for the regional analysis, in part because it includes air quality-relevant aerosol chemistry lacking in WACCM (Clayton et al., 2024). I would strongly recommend that the authors perform a detailed evaluation of a) WACCM's ability to represent baseline air quality in the present day, b) WACCM's ability to reproduce already-understood effects of climate change on air quality, and c) the likely gaps in WACCM's representation of processes and species which are important to air quality, beyond the question of tropospheric photolysis below.

We agree that the choice of model is central to the credibility of our conclusions, and we appreciate the opportunity to clarify both the strengths and limitations of CESM2-WACCM6 for air quality analysis.

First, regarding model suitability, we chose to analyse the CESM2-WACCM6 ARISE-SAI simulation output because its fully interactive chemistry–climate framework allows us to capture the coupled response of meteorology, radiation, and atmospheric composition under SAI. ARISE-SAI simulations have already been widely applied to study climate impacts of SAI, and we believe expanding its use to study air quality is a necessary next step. Although not without limitations, CESM's integration of climate, radiation, chemistry, and aerosols provides a unique opportunity to evaluate potential on-the-ground impacts of SAI. This coupling capability is essential, as air quality responses to SAI are strongly mediated by dynamical and chemical feedbacks that cannot be represented in offline or regionally constrained models. While not perfect, our analysis offers a valuable foundation for more comprehensive assessments of air quality and SAI. Moreover, CESM2-WACCM6 has previously been evaluated against observations of ozone, aerosols, and precursors. Emmons et al. (2020) provide a systematic assessment of CAM6-chem and WACCM6, showing that the model captures large-scale distributions of tropospheric ozone and key pollutants. Similarly, Gettelman et al. (2019) evaluate WACCM6's baseline climatology and variability, finding that its representation of ozone, aerosols, and chemistry is consistent with other state-of-the-art Earth system models. Griffiths et al. (2021) benchmark CMIP6 models (including WACCM) against observed tropospheric ozone distributions, demonstrating broad agreement in spatial and seasonal variability. Additional studies highlight the implementation of new chemistry mechanisms: Tilmes et al. (2019), Emmons et al. (2019), and Schwantes et al. (2020) evaluated CESM2(WACCM6) against both earlier model versions and multiple observational datasets—including NASA's ATom aircraft campaign, TOAR surface ozone, and MOPITT carbon monoxide—reporting good agreement with ozonesondes and seasonal ozone cycles, though with some regional spatial biases. For aerosols, Hancock et al. (2023) used CESM2-WACCM6 to analyze PM2.5 over India, finding that while the model underestimates concentrations in some regions due to missing species, it reproduces observed spatial patterns and long-term trends.

Regarding WACCM's ability to reproduce effects of climate change on air quality, WACCM has also been applied to evaluate well-established climate–air quality interactions. Fiore et al. (2022) demonstrate that WACCM6 reproduces observed large-scale tropospheric ozone changes in response to climate variability and anthropogenic forcing. Griffiths et al. (2021) similarly show that WACCM captures the long-term evolution of tropospheric ozone, consistent with our understanding of emissions and climate drivers.

At the same time, important limitations must be acknowledged. Regarding Hancock et al. (2023), we appreciate the reviewer raising this point. Indeed, CESM2-WACCM6 does not include explicit ammonium or nitrate aerosol chemistry, and these species can be important contributors to fine particulate matter in certain regions. Hancock et al. (2023) evaluated the performance of WACCM6 using observations of monthly PM2.5 over India and found that the model underestimates PM2.5 for certain seasons and cities due to the omission of coarse

particles, such as nitrate and ammonium, which are important components of PM2.5 in India. Ren et al. (2024) show that many CMIP6 models, including WACCM, underestimate PM2.5 burdens globally due to this omission.

Despite these limitations, Hancock et al. (2023) evaluated the spatial pattern and trends of PM2.5 and meteorological variables and concluded that air pollutant emissions, rather than climate variability, play a dominant role in poor air quality in India. In our analysis, air quality impacts are based on changes in PM2.5 and ozone. PM2.5 in WACCM is composed of six species (sulfate, organic carbon, black carbon, sea salt, and dust), so while ammonium is not included, the model still captures the major contributors to global PM2.5, and it is unlikely that sulfate-related changes would greatly affect ammonia contributions to PM2.5, making our conclusions robust to this shortcoming. That said, we recognize that ammonium and nitrate could add to the PM2.5 burden, especially in ammonia-rich regions, and this omission may lead us to underestimate pollution-related health impacts. We will make these limitations clearer in the revised manuscript (L 86).

The second concern is related. Specifically, the fact that WACCM uses fixed tropospheric photolysis rates is a significant shortcoming in a study which seeks to understand the atmospheric composition implications of stratospheric aerosol injection. This is a difficult issue to rectify, and I am glad to see that the authors have at least acknowledged this challenge. However, previous studies (e.g. Xia et al., 2017) did include this response and discussed at length the potential for tropospheric UV changes to be significant in understanding the tropospheric ozone response – and thus the air quality response. The authors themselves argue that tropospheric photochemistry is the dominant factor in NH surface ozone change (line 219). Ideally, an analysis such as that by Clayton et al. (2024) in which WACCM outputs are used as boundary conditions to a more air quality-focused model may be a way to resolve these issues, and I would recommend that the authors seriously consider if there is a way that they could perform a more comprehensive simulation of tropospheric chemistry using their existing data - recognizing that this would require a great deal of additional work but would also resolve what I perceive as being a major gap in the work.

We realized that stating that the model does not include photolysis changes is incorrect. The model includes photolysis rates that are calculated using lookup tables accounting for overhead ozone column and clouds. However, this approach of calculating the photolysis rates does not include the direct radiative effects of dynamic aerosol distributions. In an attempt to address this gap, we conducted offline calculations with the Trospheric Ultraviolet and Visible (TUV) model, which indicate that surface UV changes under SAI are small and broadly consistent with past studies (Bardeen et al., 2021).

Our analysis focuses on the health impacts of particulate matter (PM2.5), which is the primary driver of air pollution–related mortality worldwide. While we acknowledge that our framework likely underestimates the role of tropospheric photochemistry in shaping ozone changes, the main conclusions regarding PM2.5 impacts remain robust. We have highlighted this limitation in

the text and interpreted the ozone-related results with appropriate caution (L 380). We view our study as a first step in quantifying global-scale mortality implications of SAI and hope that future work can build upon it by explicitly incorporating variable tropospheric photolysis rates and extending the analysis to UV-driven health outcomes (e.g., skin cancer, cataracts) that lie outside the scope of the present paper.

Notwithstanding such an expansion, these are sufficiently significant deficiencies that I believe they need to be much more strongly highlighted. I would recommend that the abstract explicitly state that changes in tropospheric photolysis are not considered, and that statements that this is the first study to use "comprehensive" stratospheric and tropospheric chemistry (e.g. line 29 and 58) be removed. While I absolutely believe that this study can provide a valuable contribution to our understanding of the impacts of SAI on the environment, I would argue that it needs to be placed in the correct context (and thus allow subsequent studies to fill the remaining knowledge gaps).

We have revised the manuscript to remove the phrasing of the model as "comprehensive" and avoid suggesting that this is the first study with fully comprehensive stratospheric and tropospheric chemistry. In the abstract, we have added explicit caveats about photolysis/UV treatment and clarified that the results are scenario-specific, not general for all SAI.

Independent of these concerns, I was struck by one of the conclusions drawn (and which is highlighted in the abstract). The authors argue that internal variability is key, on the basis that they find significant differences across ensemble members. This aligns climate intervention effects on air quality with the well established effects of climate change on air quality (e.g. Fiore et al., 2015) where noise in the meteorological response can be greater than the change in exposure to pollutants resulting from SAI. It would have been useful to discuss how the projected effects of SAI on air quality compare to the air quality "penalty" projected for climate change, given that there is a robust literature discussing not only this question but also specifically the problem of how to deal with internal variability in such projections. The lack of such a discussion is a notable absence, and leaves the paper somewhat unmoored.

We thank the reviewer for highlighting the importance of exploring the parallel between internal variability in climate intervention studies and the air quality "penalty" literature. We agree that this is an important contextual point, and we have added a new discussion at the end of the conclusion (see page 18, last paragraph) comparing our findings to prior work.

The use of large ensembles is a good (if expensive) solution to this problem, but analysis of air quality interventions may also rely on representative meteorological years if it can be shown that the outcome would be the same as when using a large ensemble average (see e.g. Stewart et al. (2017) and Abel et al. (2018) for examples looking at air quality change in future conditions). Here it seems that internal variability is used to draw some conclusions which seem hard to justify; for example, on lines 291-293 it is claimed that "health impacts under SAI are not governed mainly by the magnitude of SO2 injected".

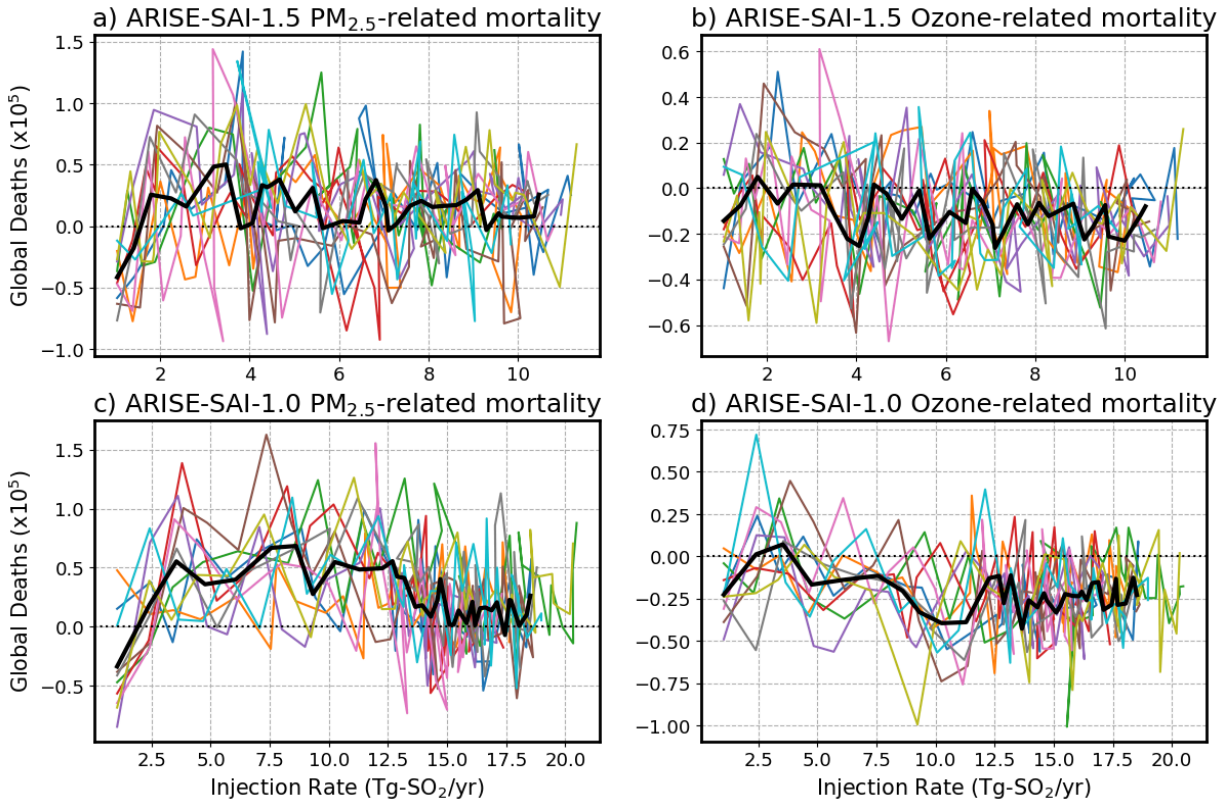
Certainly it is true that SAI alone is not going to become the dominant cause of air pollution under almost any scenario, and the comparison of ARISE-SAI-1.0 and 1.5 shows how important these other factors are - an important contribution. However the paper simultaneously argues that there is a robust surface ozone response relative to a scenario where the amount of SO2 injected is zero (SSP2-4.5), so presumably the magnitude of the injection is not entirely irrelevant. Is there evidence that a robust (if complex) difference in the effects of larger injection quantities would not emerge if using a larger ensemble, longer averaging period, and/or if other factors (eg surface-level emissions of air quality precursors) were held constant? I would suggest that the authors explore in more detail the degree to which their results might be improved by such approaches, not least because the data to do so appears to already exist (e.g. it should be straightforward to evaluate the degree to which a smaller ensemble would or would not have allowed the same conclusions to be drawn - which would be valuable information for those interested in performing future studies of atmospheric composition change under SAI).

We thank the reviewer for raising this important point. In Figs. 7 and 8, we show that air pollution-related mortality does not increase monotonically with the magnitude of SO2 injected across ARISE-SAI-1.0 and ARISE-SAI-1.5. There is a robust surface ozone response, which likely arises from SO2 being injected primarily in the southern hemisphere during 2060-2069, that modulates the hemispheric temperature gradients. For surface ozone, the impact on mortality is a complex interplay between deposition, tropospheric and stratospheric changes due to chemistry and transport, and surface changes due to cooling. However, our results indicate that such an interplay results in a significant, but not magnitude-dependent, change. Furthermore, changes in surface ozone impact the spatial distribution of ozone-related mortality but not the global average.

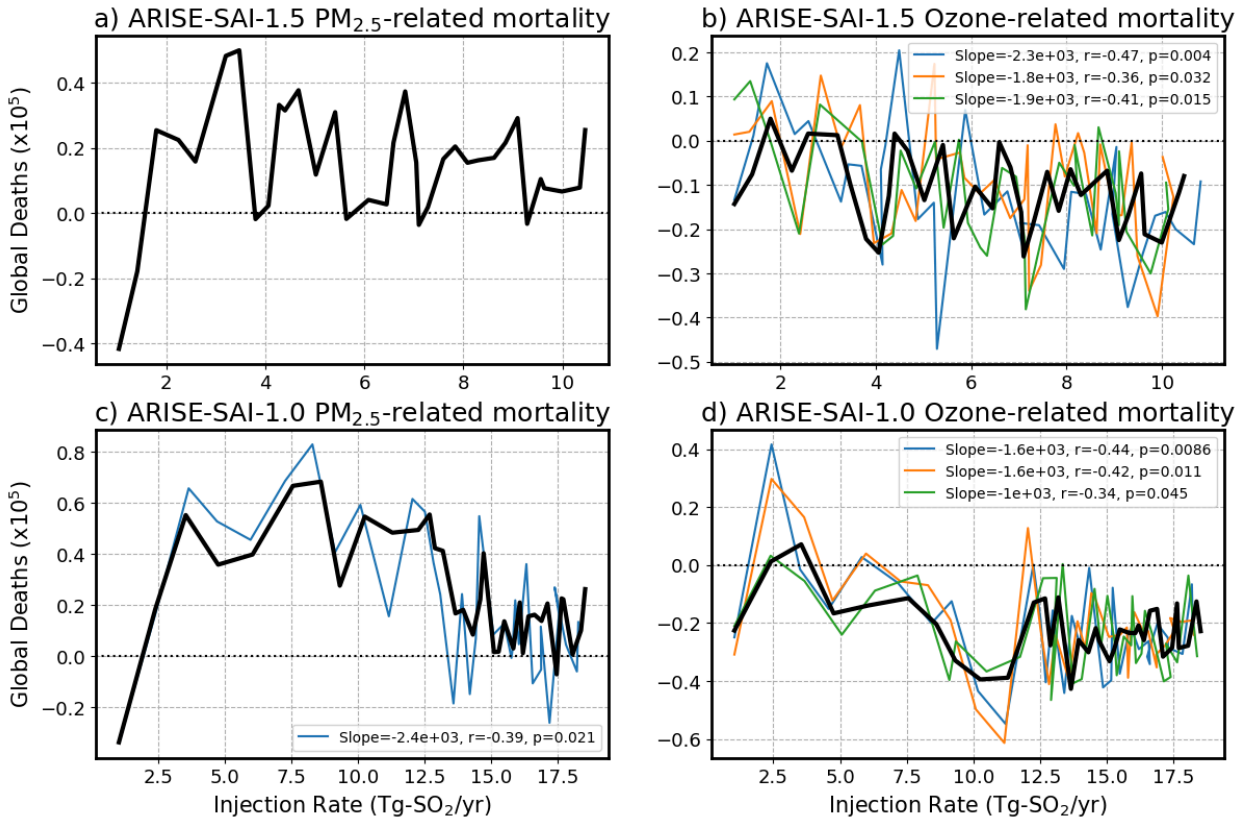
In additional simulations following the G2-SAI-3DOF and G2-SAI-1DOF protocols (Visioni et al., 2024), where injections occur *without* any associated changes in tropospheric chemistry, we still find a robust surface ozone response relative to SSP2-4.5.

Regarding ensemble size, our analysis draws on two independent 10-member ensembles (ARISE-SAI-1.5 and ARISE-SAI-1.0). A 10-member ensemble is generally considered sufficient to separate forced responses from internal variability (Milinski et al., 2020; Wills et al., 2020), and indeed, our results show consistent outcomes across both ensembles. Furthermore, by analyzing 35-year time series (see figure below), we find that the overall mortality estimates do not exhibit a clear trend with increasing injection amounts, regardless of the temporal averaging resolution. Therefore, having a large ensemble size gives us more confidence in our results. However, we agree that using representative meteorological years or expanding the ensemble size further could provide additional insights, and we have revised the text to explicitly acknowledge this (last sentence of Section 3.2). To address this concern, we performed a subsampling analysis: drawing 10 random sets of 3 ensemble members each, we found that only a minority of subsamples exhibited a statistically significant linear relationship ($p < 0.05$) between injection rate and mortality, and even then, the correlation coefficients were weak. This

suggests that our central conclusion that mortality impacts do not scale in a simple linear fashion with SO2 injection remains robust, though further ensembles, longer time periods, and controlled precursor emissions experiments would be valuable in future work. Both figures have been added to the supplemental.



Global differences in air-pollution mortality between ARISE-SAI and SSP2-4.5 as a function of the total SO2 injection rate for the first 35 simulation years. Panels show (a) PM2.5-attributable mortality for ARISE-SAI-1.5, (b) ozone-attributable mortality for ARISE-SAI-1.5, (c) PM2.5 for ARISE-SAI-1.0, and (d) ozone for ARISE-SAI-1.0. Colored lines indicate individual ensemble members, while the thick black line represents the 10-member ensemble mean.



Global mortality differences between ARISE-SAI and SSP2-4.5 are shown as a function of the SO2 injection rate over the first 35 years of simulation. Panels depict (a) PM2.5-attributable mortality for ARISE-SAI-1.5, (b) ozone-attributable mortality for ARISE-SAI-1.5, (c) PM2.5 for ARISE-SAI-1.0, and (d) ozone for ARISE-SAI-1.0. Colored lines represent ordinary least-squares fits from randomly selected 3-member ensemble subsets (10 draws), plotted only where significant linear relationships are detected ($p < 0.05$). Slopes, r-values, and p-values are annotated in the legend.

Minor comments

Some aspects of the air quality response which I had expected might be significant were seemingly not discussed. I would recommend discussing whether elements of the air quality response to SAI which have been significant for studies of the climate penalty – for example, changes in planetary boundary layer height, and the (highly model-dependent) lightning response – are playing a significant role in the calculated response. These factors are well described in the literature already cited and would be expected to be represented in an ESM (ostensibly one of the key novelties of this work), so providing a careful evaluation of how these factors translate to an SAI study would be valuable.

We appreciate the reviewer's suggestion and agree that factors such as planetary boundary layer (PBL) height and lightning are important components of the broader air quality–climate

literature. However, the scope of the present study is intentionally focused on quantifying the air quality and associated health impacts of SAI, using CESM to evaluate the net surface-level changes in PM2.5 and ozone concentrations across large ensembles. Our goal here is not to provide a mechanistic attribution of every pathway by which SAI may influence surface air quality, but rather to assess the aggregate outcome of these multiple processes as represented within the model.

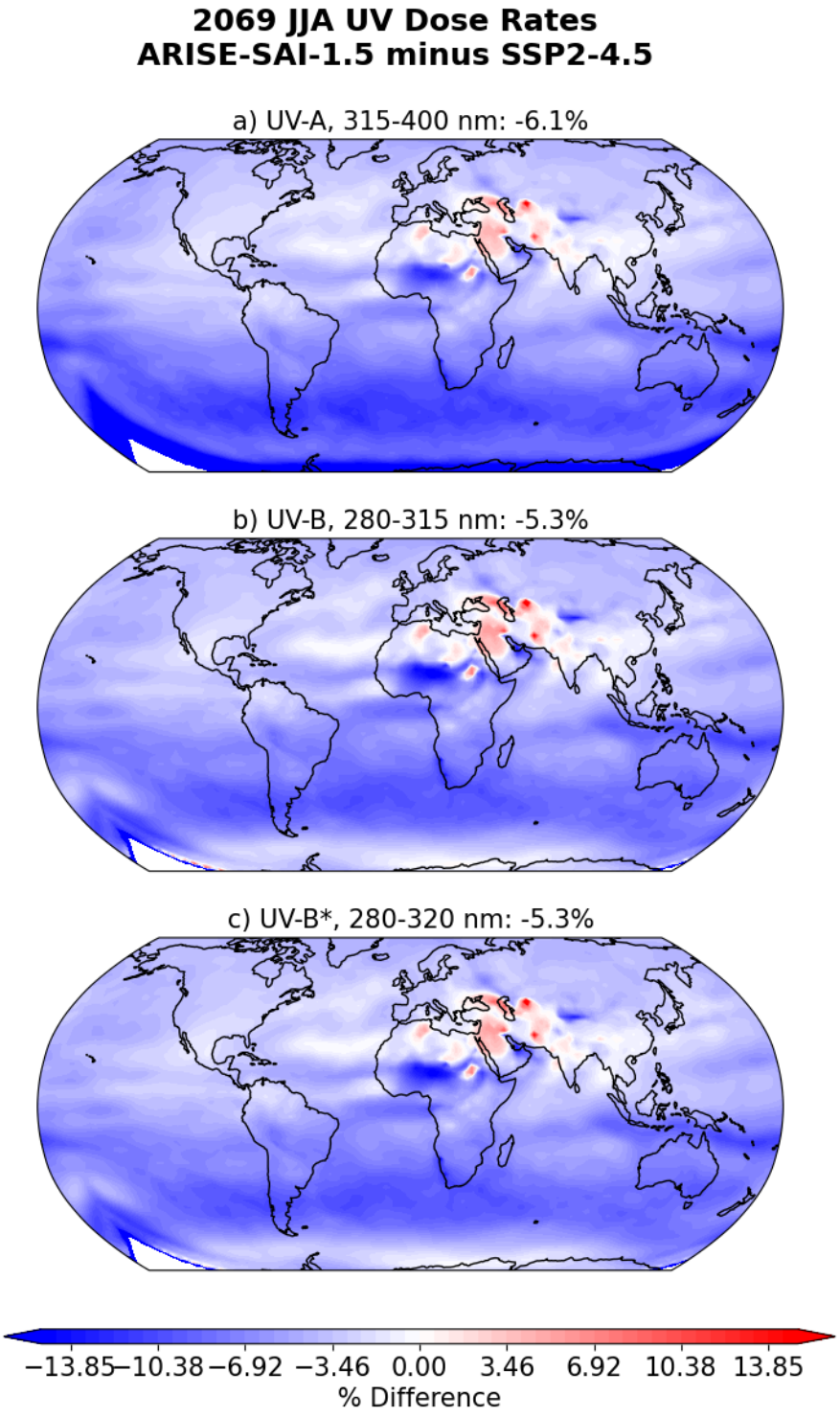
Elements such as PBL height and lightning response are indeed simulated in WACCM and therefore implicitly contribute to the overall modeled response. To address the reviewer's comments, we have included a plot of PBLH changes in the supplemental which show interesting changes in PBLH from SAI but a more detailed process-level analysis of each of these mechanisms is an important and valuable direction for future work. We frame our analysis around the ensemble-mean concentration and mortality responses, and we highlight where internal variability and policy-driven changes dominate the signal. By design, this allows us to place the air quality consequences of SAI in direct context with prior studies of the climate penalty and emissions controls, while keeping the focus on the net implications for surface air quality and health outcomes. We have clarified this point in the text and noted that more detailed process studies, which include explicit evaluation of changes in PBL dynamics and lightning, will be a valuable complement to our findings (L 385).

While I understand why the authors have chosen not to estimate the health impacts of UV changes associated with SAI, I was surprised that no formal analysis was done at all of surface UV changes. The statement on line 372 – that a preliminary analysis indicated "very modest changes" – is unfortunately not much help, as the authors do not provide any metric of what they consider to be "modest" (or why). Quantifying (say) relative changes in projected population exposure to UV would help us to understand whether such changes need further study. Quantitative analysis of UV changes may also be useful in understanding the degree to which neglecting changes in tropospheric photolysis change may or may not be a minor oversight.

We agree that quantifying surface UV changes provides useful context for understanding both health and chemical implications. For the preliminary analysis in question, we conducted an offline analysis of the UV changes using the Tropospheric Ultraviolet and Visible Radiation Model (TUV-X; https://github.com/NCAR/tuv-x). We calculated photolysis rate constants and surface UV changes under clear-sky conditions, comparing output from the two simulations. Our analysis indicates that percentage changes in surface UV are between -5.3 to -6.1% globally. For example, for JJA 2069 we find relative changes on the order of only a few percent between the SAI and SSP2-4.5 scenarios.

In response to this comment, we have added a supplemental figure (shown below) showing the spatial distribution of percentage changes in surface UV for JJA 2069. This figure illustrates that changes are small across nearly all regions. While this analysis confirms that UV changes are not a dominant driver of the air quality responses we focus on here, we agree that they remain relevant for future work, especially in the context of quantifying potential UV-related health

effects. The text has been revised to include a discussion of our preliminary findings from TUV and how it is consistent with previous work (L 391).



Hancock et al. (2023) indicated that WACCM-based estimates of exposure to PM2.5 may overestimate the role of dust, due to inclusion of too-large particles in the PM2.5 metric.

Given that dust is the predominant factor in exposure under ARISE-SAI-1.5 for a significant fraction of the world (Figure 2), it would be useful to have more information on how the PM2.5 calculation was performed and whether the issue identified by Hancock et al. was addressed.

The reviewer raises an important point. Our PM2.5 calculation follows the same setup described in Hancock et al. (2023), and it is therefore subject to the same caveats regarding the representation of dust, including the potential inclusion of overly large particles in the PM2.5 metric. We acknowledge that this may lead to some overestimation of dust contributions to total exposure. However, we emphasize that our analysis focuses on differences between the ARISE-SAI-1.5 and SSP2-4.5 scenarios, rather than the absolute magnitudes of exposure. Because the same definition of PM2.5 is applied consistently across both scenarios, any systematic bias in the representation of dust is expected to cancel out when examining the relative effects of SAI. For this reason, while the caution identified by Hancock et al. is relevant to the interpretation of the absolute dust burden, it does not materially affect the conclusions we draw about the differences attributable to SAI. We have included text in the manuscript to highlight this caveat.

There are numerous grammatical errors (e.g. lines 221-222: "many of this conditions", "we deem important"; line 228: "These estimates and Fig. 4 show that the standard deviation of mortality estimates highlights the large spread in project PM2.5-related deaths"; Eq. 2 says the PM2.5 threshold is 2.4 (no units given), but Table 1 says 2.5 ppm - and Burnett et al (2018) say 2.4 ug/m3). I would recommend the authors take some time to go through the paper in depth and fix such errors before resubmitting.

The text has been revised to address these grammatical errors.

Bibliography

Bardeen, C. G., Kinnison, D. E., Toon, O. B., Mills, M. J., Vitt, F., Xia, L., ... & Robock, A. (2021). Extreme ozone loss following nuclear war results in enhanced surface ultraviolet radiation. Journal of Geophysical Research: Atmospheres, 126(18), e2021JD035079.

Emmons, L. K., et al. (2020). The Chemistry Mechanism in the Community Earth System Model Version 2 (CESM2). Journal of Advances in Modeling Earth Systems, 12(4), e2019MS001882. https://doi.org/10.1029/2019MS001882

Fiore, A. M., Hancock, S. E., Lamarque, J. F., Correa, G. P., Chang, K. L., Ru, M., ... & Ziemke, J. R. (2022). Understanding recent tropospheric ozone trends in the context of large internal variability: a new perspective from chemistry-climate model ensembles. Environmental Research: Climate, 1(2), 025008.

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., ... & Randel, W. J. (2019). The whole atmosphere community climate model version 6 (WACCM6). Journal of Geophysical Research: Atmospheres, 124(23), 12380-12403.

Griffiths, P. T., Murray, L. T., Zeng, G., Shin, Y. M., Abraham, N. L., Archibald, A. T., ... & Zanis, P. (2021). Tropospheric ozone in CMIP6 simulations. Atmospheric Chemistry and Physics, 21(5), 4187-4218.

Hancock, S., Fiore, A. M., Westervelt, D. M., Correa, G., Lamarque, J. F., Venkataraman, C., & Sharma, A. (2023). Changing PM2. 5 and related meteorology over India from 1950–2014: a new perspective from a chemistry-climate model ensemble. Environmental Research: Climate, 2(1), 015003.

Milinski, S., Maher, N., & Olonscheck, D. (2020). How large does a large ensemble need to be?. Earth System Dynamics, 11(4), 885-901.

Nolte, Christopher G., et al. The potential effects of climate change on air quality across the conterminous US at 2030 under three Representative Concentration Pathways. Atmospheric Chemistry and Physics. 2018

Ren, F., Lin, J., Xu, C., Adeniran, J. A., Wang, J., Martin, R. V., ... & Takemura, T. (2024). Evaluation of CMIP6 model simulations of PM 2.5 and its components over China. Geoscientific Model Development, 17(12), 4821-4836.

Schwantes, R. H., Emmons, L. K., Orlando, J. J., Barth, M. C., Tyndall, G. S., Hall, S. R., ... & Bui, T. P. V. (2020). Comprehensive isoprene and terpene gas-phase chemistry improves simulated surface ozone in the southeastern US. Atmospheric Chemistry and Physics, 20(6), 3739-3776.

Tilmes, S., Hodzic, A., Emmons, L. K., Mills, M. J., Gettelman, A., Kinnison, D. E., ... & Liu, X. (2019). Climate forcing and trends of organic aerosols in the Community Earth System Model (CESM2). Journal of Advances in Modeling Earth Systems, 11(12), 4323-4351.

Wills, R. C., Sippel, S., & Barnes, E. A. (2020). Separating forced and unforced components of climate change: the utility of pattern recognition methods in large ensembles and observations. Variations, 18(2), 1-10.