Reviewer comments are in bold. Authors' responses are in blue.
Response to reviewer #2

This paper reported one set of the experiments in the GeoMIP – G6solar and G6sulfur using 6 climate model output. Although models successfully reach temperature target (from the level of SSP5-8.5 to the level of SSP2-4.5), there are different climate responses between G6solar and G6sulfur, and large inter-model spreads in many climate aspects. The manuscript raises lots of questions on model uncertainties and following-up impact studies. In general, this study is strong and important as it is the first thorough report on G6solar and G6sulfur from 6 models, and it indicates many future research directions for future. But the manuscript needs some improvements in the writing style before publishing. The author tends to use long and obscure sentences when describing the figures and explaining the underline hypothesizes.

We thank the reviewer for their encouraging comments, and for their help in improving the clarity of our manuscript, which is really appreciated. We respond to all their points below.

Specific comments:
Line 7-9: it is better to change to something like "We find that, over the two decades of the century, there are considerable inter-model spreads in the needed injection amounts of sulfate, in the latitudinal distribution of the aerosol cloud, and in the stratospheric temperature changes resulting from the extra aerosol layer."

Thank you for the suggestion. We have changed the phrase accordingly.

Line 12-13: Are those values averaged differences between SSP5-8.5 and SSP2-4.5 over 2081-2100? If so, it is better to make it easy to read. Something like "SSP5-8.5 minus SSP2-4.5 averaged over 2081-2100"

They are. We have changed the phrase.

Line 13: please change ": the differences in the simulated aerosol spread then change some of the underlying uncertainty, for example in terms of" to ". With aerosol injection, the differences of aerosol spread further change some of the underling uncertainties, such as"

Changed.

Line 16: please change "a larger inter-model spread in the regional response in the surface temperatures" to "a larger uncertainty in the regional surface temperature response among models"

Done. Thanks for the suggestion.

Line 44: please clarify "with no baseline simulation to analyse the response against (as in the case of G4)".

We understand this was unclear. We changed that to specify we meant that "*In the case of the G4 experiment, furthermore, there was no sustained future scenario with similar global surface temperatures achieved without geoengineering, but with less CO_2, to compare the results against*"

Line 54: "Both reductions" of what? The temperature reductions are checked every decade, and then the sulfate injection amount and solar radiation reduction are adjusted.

We added the words "of incoming surface insolation" to clarify.

Line 55: please reorganize this sentence "There are multiple uncertainties…intercomparison"

Done!

Line 62: add ")" after "Visioni et al. (2017b)"

Added.

Line 64: add "SO2" after "Tg"

Done.

Table 1: The first column – model names, are hard to read. Maybe add one extra space among models?

We have rewritten the models' name in bold for clarity.

Line 99: CESM2 also injects SO2 between 10N-10S? Or following the feedback algorithm and injecting SO2 from other latitudes?

As a mistake, CESM2 only injected at 0°N.

Line 100: how could CNRM-ESM2-1 use SO2 distribution file from G4SSA for G6 experiment?

They used the aerosol distribution, not SO2. We have changed the text for clarity.

Line 102: change ";" to "and"

Fixed.

Figure 1: Please keep all sub title styles consistent.

Done!

Line n 116-120: please reorganize this sentence. It is too long.

We have split the sentence up.

Table 2: average of which period? In the title, it is said "the last decade of the experiment", and also said "2081-2100". Please use "SSP58.5 minus SSP24.5" instead of "SSP58.5-SSP24.5". And please use minus instead "-" in the whole manuscript.

We have specified it's the last *two* decades, and tried to change the text according to the reviewer's suggestion when possible (except in the table, for reasons of space).

Figure 2: R2 in d is zero? What is m?

Yes, it's 0. We have updated the caption for clarity (m is the slope of the linear fit).

Line 131-132: how are those numbers calculated? What does this sentence mean?

Following the request from reviewer 1, we have included in the supplementary the non-normalized version. We have updated the text for clarity.

Line 144: please indicate which panels a and b are.

Done!

Line 145: why the comparison only helps the first 30 years?

Because some models show very little changes in TOARF in that period (for instance, CESM2) and that explains why the amount of intervention is very small, or constant, in there.

Line 148: please clarify the sentence. As far as I understand, the first part of the sentence means that the small differences of global mean temperatures anomaly between SSP5-8.5 and SSP2-4.5 among models tend to magnify the inter-model differences of intervention applied. But what does "resulting in larger differences in the first years" mean?

We have tried to clarify. We mean that, with small forcings, the estimates made by the single modeling teams would have magnified some of the differences.

Line 148, 150: please change "first decades" and "first years" to "first couple decades" and "first several years". Or indicate exact numbers.

Done.

Line 145-153: I still don't understand the different mechanisms behind the two periods (first three decades and the rest)

We hope the clarifications in the text have helped. For instance, if 0.3 W/m^2 in TOARF have to be balanced in the first decades, and the sensitivity of the estimate is around 0.1 W/m^2, the error can be large. Towards the end, if the imbalance is 2.0, the same sensitivity influences less the results.

Line 161: "different models' behavior"

Fixed.

Line 164: "reasons for a different aerosol distribution with similar injection locations and height of SO2 are"

Changed.

Figure 4: CNRM-ESM2-1 and MPI-ESM are both prescribed SO2 distribution files. How could MPI-ESM have SO2 injection amount, but CNRM-ESM2-1 not in b)? The box is hard to read. It is better to draw it a table. If all three panels use the same color code, then only one legend is needed.

MPI derived their aerosol distribution from a similar version of the model without interactive ocean or land (to save computational time): it is therefore consistent to determine the SO_2 amount from the scaling of the required AOD: CNRM uses another model's distribution, therefore it wouldn't be. We fixed the figure as suggested.

Line 183: please clarify this sentence. "Model spread, … observations)."

Uncertainty is related to the distance between simulations and real world (also known as accuracy). Models' spread is not, therefore, the same as uncertainty. We have clarified the phrase changing it to "*Model spread on a particular result is not, of course, the same as uncertainty: models may agree despite a lack of observational support, resulting in a narrow spread that might be inaccurate, or the spread might be large because some model results are simply inconsistent with available observations*".

Line 209-210: are those values (e.g. CESM is 6.2 at 2058) in models accumulated SO2 or SO2 per year? Based on Fig. 4b, they are SO2 Tg/yr? If so how this annual injection amount compare to one time injection from Pinatubo? Before the AOD reaching the level of that in Pinatubo, there have been decades of injection already.

They are annual. It is, of course, a rough comparison, but given a 1 year lifetime for stratospheric aerosols, they can be somewhat compared.

Line 231: UKESM1-0-LL is showing much weaker stratospheric heating, but from Figure 5, it seems that CNRM is the least? Also IPSL shows similar stratospheric heating as in MPI.

We corrected the phrase.

Line 251: "Banerjee et al. (2020))"

Fixed.

Line 259-261: please reorganize this sentence "not only are …. Removed"

Done

Line 261: "For G6sulfur, there is a general model agreement in the …"

Changed.

Line 264: “as evidenced by the absence of high-latitude warming with the same magnitude in the G6solar simulations”

Done.

Line 275: please reorganize the first sentence.

Changed.

Line 277: why saying “that is balanced in the multi-model average by a stronger warming modeled by MPI-ESM”? all models are showing the warming over Northern Eurasia with different magnitudes. There is nothing to be balanced.

The reviewer is right. We removed the phrase.

Line 298-299: Both versions of MIP show the warming in North America and West Antarctica, just one is much stronger than the other.

Fixed.

Line 299-300: How could the different responses in HR and LR indicate that the deep ocean circulation is causing regional differences in temperature? What is the difference between HR and LR?

The difference is just in the horizontal resolution, so this is just our assumption. We hope to investigate this further in the future!

Line 336 to 344: CESM seems showing the least difference between G6sulfur and G6solar. Actually they are almost the same amount of reduction.

Changes in precipitation are much larger East of Australia, for CESM2, however. We added a note anyway.

Line 345: please reorganize this sentence to something like “Lastly, models agree on regional precipitation changes more in G6solar than in G6sulfur”

Done!

Figure 10: legends cover part of the plots.

To do

Line 351-352: HR shows stronger reduction and increasing of precipitation in the tropics relative to LR, not just stronger reduction.

Fixed.

Line 354: Again, why are circulations different in HR and LR?

We added a note specifying that the horizontal resolution is different (thus we can assume the circulation behaves differently).

Line 358-363: please reorganize this sentence.

Done!

Figure 11: Hard to read. Some suggestions: 1. Order the main categories with altitude: Surface at the bottom, Stratosphere on the top, Radiative forcing and Aerosol chemistry/microphysics are through the whole troposphere and stratosphere. 2. "Stratospheric aerosols" and "Stratospheric Heating" should be in the same category. Stratospheric aerosol is what we injected in the model, and all others (including stratospheric heating, dynamics and chemistry…) are responses to the extra aerosols

Thanks for the suggestions. We have revised the figure accordingly.