Response to Anonymous Referee #1

We thank the Referee 1 for sending very helpful comments and suggestions to the manuscript. All the comments are addressed below in detail:

This study seeks to propose new testbed model experiments for studying scenarios of stratospheric aerosol geoengineering (SAG) designed to limit global warming to fixed global mean surface temperature targets, with some additional constraints to limit un- desirable side-effects. I appreciate the interest that the authors' idea has for the earth system community, and I find the paper to be generally well-structured and with a clear logical flow. However, as far as I can tell the main novelties of the study lie in the use of a very recent CMIP6 model, and in the combination of a feedback controller modelling approach with overshoot climate scenarios – neither of which is novel in iso- lation. Several of the conclusions are likely highly model dependent, and the broader considerations echo the results of other recent studies. Furthermore, some aspects of the manuscript – for example the figures – have a very unrefined feeling. Finally, I have two major concerns on the structure and contents of the study, which I detail below. Based on this, even though the topic of the study is well-suited to ESD, I am not convinced that it is suitable for publication in this journal.

Based on the two referees comments we realized that the framing of the paper, on the one hand, proposing a test-bed simulation for GeoMIP and on the other hand, discussing novel findings, did not adequately convey the contributions of the paper – in particular, that the impacts of stratospheric aerosol geoengineering strongly depends on various different aspects of the experiment, the considered baseline scenarios and therefore the CO2 concentrations, the amount of SO2 injection, and the chosen temperature targets. To better emphasize the contributions, in revising the paper we will focus on the novel findings, and use those, along with the potential for model-dependent outcomes, as motivation for the suggestion that these could be new test-bed simulations for GeoMIP. The referee is correct that main novelty of the experiments lies in the fact that we have combined the feedback controller with the overshoot scenario. Other novelties, that we have failed to point out more clearly and plan to mention in the revised manuscript, include that for the first time these types of simulations use the CMIP6 future pathways, which are unique since they are based on socio-economic considerations. Furthermore, we are using an updated Earth-System model, that includes more impact relevant coupling, including interactive crop models, land-ice model, and ocean-bio-chemistry. This paper also serves as an overview paper that describes the general setup of the experiments, while additional papers that are in preparation that will refer to this paper. We also agree with the referee, that the results are highly model dependent. While some of the results are aligned with earlier findings, the focus of the paper is not on repeating what has been done before, but describing potential impact relevant outcomes and for other modeling groups to repeat these experiments in order to produce multi-model comparisons to help determine uncertainties of outcomes. One example is already provided in this paper, comparing the Atlantic Meridional Circulation changes with earlier model results. Further, we have been improving the figures to be added the revised version of the manuscript as suggested by the referee.

The abstract will be modified to support the above points and the main text will be modified accordingly:

New Abstract: "A new set of stratospheric aerosol geoengineering (SAG) model experiments have been performed with CESM2(WACCM6) that are based on the CMIP6 overshoot scenario (SSP5-34-OS) as a baseline scenario to limit global warming to 1.5°C or 2.0°C above 1850-1900 conditions. A feedback algorithm has been used to identify the needed amount of sulfur dioxide injections in the stratosphere at four predefined latitudes, 30°N, 15°N, 15°S, and 30°S, to reach three surface temperature targets: global mean temperature, and inter-hemispheric and pole-to-equator temperature gradients. The combination of using an overshoot scenario as a baseline that limits the needed amount of SAG applications and the use of a feedback algorithm to reach pre-defined temperature targets in model experiments is expected to reduce some of the earlier identified side effects of SAG. These experiments are therefore relevant for investigating the impacts on society and ecosystems. Comparisons to SAG simulations based on a high emission pathway baseline scenario (SSP5-85) further help investigate the dependency of impacts using different injection amounts to offset surface warming by SAG. We find that changes from present day conditions (2015-2025) in some variables depend strongly on the defined temperature target (1.5°C vs 2.0°C). These include surface air temperature and related impacts, the Atlantic Meridional Overturning Circulation (AMOC), which impacts ocean net primary productivity, and changes in ice sheet surface mass balance, which impacts sea-level rise. Others, including global precipitation changes and the recovery of the Antarctic ozone hole, depend strongly on the amount of SAG application. Furthermore, land net primary productivity as well as ocean acidification depend mostly on the global atmospheric CO$_2$ concentration and therefore the baseline scenario. Multi-model comparisons of the experiments proposed here would help identify consequences of scenarios that include strong mitigation, carbon dioxide removal with some SAG application, on societal impacts and ecosystems."

Major Comments

1. The study performs only one simulation for each geoengineering experiment, citing computational limitations as the main reason. Since the authors state that the study's goal is to establish a protocol for new model experiments, this is justifiable. However, the authors then perform only three SAG experiments; the obvious absent is Geo-SSP85 2.0. Given that comparing SAG interventions with the same temperature goals under different scenarios is a major focus of the study, and that – as the authors themselves underscore – past simulations with earlier model versions show significant differences from the ones presented in this study, I struggle to see the logic in not including such a simulation.

We have now finalized a second ensemble member for each of the simulations presented in the manuscript, and we believe that our findings are more robust with those. Our conclusions in this paper that outcomes of geoengineering strongly depend on different baseline scenarios, injection amount and target temperatures, are supported with the simulations we have presented. Based on the available simulations Geo-SSP5-34-OS 1.5 and Geo-SSP534-OS-2.0, we are able to discuss differences between using different target temperatures. Since computer time is a main issue, we decided to add a second ensemble to the existing experiments, because we think that we can stronger support our conclusions that we could gain from the existing experiments, and not add the additional 2.0 experiment to the high forcing scenario.

For the testbed experiment, we decided to focus on the overshoot scenario. We will explain more clearly that the high forcing scenario is not only performed for comparisons with a high forcing scenario but also to be able to identify difference to the earlier study, using a different model version. In the revised manuscript, we will clarify that the proposed experiments are based on the Geo-SSP5-34-OS baseline scenarios a not on the high forcing scenarios, because this is a more relevant scenario for impact analysis. The overshoot experiment does not require unsustainable amounts of SO2 injections, which provide a potentially more policy-relevant scenario.

2. The study reads as a generally well-structured, primarily descriptive report of a set of three SAG model simulations. If the aim of the study is indeed to describe new numerical simulations, then I would expect to see a larger number of different experiments, ensembles etc. If, instead, the goal is to establish a protocol for new model experiments, I would expect significant additional analyses and tests on the feedback controller, the latitudes of injection of the aerosols etc. The study is therefore in a grey area between a description of new numerical simulations and a more technical/mechanistic experiment design work, and I find it somewhat unsatisfactory under both categories.

We thank the referee for pointing this out. As discussed above, we will shift the focus of the paper to describing and discussing the new numerical simulations and performed a second ensemble member to this study. Based on the findings, we are still planning to recommend that it would be beneficial if the experiments based on the overshoot scenario are performed by other modeling groups to identify the range of outcomes of impact relevant diagnostics. The referee suggests to establish a protocol for new model experiment, additional analysis and tests of the feedback controller are required. We have defined specific of the experiment, including injections at four fixed altitudes at 5 km about the tropopause, and using a feedback controller that will check annual deviations for the defined temperature goals. We do not suggest that modeling groups are using a different setup, since this would complicate analysis in comparing outcomes of different models. However, we agree to provide a better description on details on the implementation of the feedback controller in the revised version of the manuscript as addressed below.

3. I find the figures unsuitable for publication. Some examples: the styles differ across figures and panels within the same figure (e.g. Fig. 4, top row vs. middle and bottom rows), colour-coding/labelling of experiments is inconsistent (e.g. cf. Fig. 1, 3, 11), some figures have panel labels (e.g. Fig. 9), while others do not, different map projections are used (e.g. cf. Figs. 5, 8 and 9) etc. I provide an incomplete list of suggestions in the minor comments below.

We agree with the referee that the figures can be much improved and will apply the same style within figures and use consistent labelling and map projections (including color coding) in the revised version of the manuscript. Changes to the figures pointed out by the referee are shown below. The remaining figures will be completed for the revised version of the manuscript.

Some Minor Comments

1. Introduction: keeping in mind the relatively broad readership of ESD, it would be useful to add one or two sentences explaining what a "feedback controller" is in this context.

We agree with the referee and will add more information regarding the feedback controller (both briefly in the introduction, and in more detail in section 2 as suggested below). For the introduction, we will add: "**In each year of the simulation, the amount of injection to use at each latitude was adjusted based on the deviations in meeting these goals [the global average surface temperature, as well as the equator-to-pole and interhemispheric temperature gradients, noted in the previous sentence]. In this way, the appropriate injections to use to meet the goals was "learned" as the simulation ran, compensating for uncertainty and avoiding lengthy trial runs**.".

2. Sect. 2.2: echoing the above comment, the description of the feedback control algorithm in this section is poor. Please rephrase and expand it. A practical example of its functioning would be beneficial.

We agree with the referee and will add more information on the feedback controller to the revised version of the manuscript so modeling groups can repeat the experiment.

3. p. 2 ll. 44-46 This is a somewhat awkward sentence, please rephrase.

In the revised version of the manuscript we will rephrase the text to:

"GLENS was based on a high forcing future climate scenario (RCP8.5) and required an increasing amount of sulfur injection with time. GLENS simulations have shown that **reaching** global surface temperature and temperature gradient **targets,** results in benefits with respect to temperature related impacts compared to experiments that only **control** for global surface temperature (Kravitz et al., 2019)."

4. p. 6 l. 164 algorithem –> algorithm

We will fix this in the revised version of the manuscript

5. Sect. 3 The authors use the term "efficiency" in the title, but never refer back to this in the section. I would suggest either discussing this in the text or removing the term altogether.

Efficiency in Section 3 has been discussed based on Figure 3, bottom panel. We will clarify this in more detail in the revised version of the manuscript

6. p. 6 l. 184 tropospheric –> troposphere.

Agreed.

7. p. 8 l. 222 "For the baseline simulations, temperatures in high latitudes are higher than in mid and low latitudes" Perhaps the authors mean "temperature anomalies"?

Thanks, we will correct this.

8. p.8 l. 230 "1.5 ËŽC and 2.0 ËŽC" –> 1.5 ËŽC or 2.0 ËŽC

Thanks, we will correct this.

9. Table I: I would suggest adding a column with the models used, as I understand that these vary between the RCP and SSP simulations.

Thanks, we will add an additional column to the Table 1 and add additional numbers for the second ensemble member of each experiment.

10. All figures: add panel letters to all figures, which makes referencing more straight forward and concise (avoiding sentences like: "Fig. 4, middle and bottom panels on the right").

Thanks, we will add letters to the figures.

11. Fig. 1 In the top panel there seems to be a large gap between the SSP scenario and the beginning of the Geo SSP5-34-OS 2.0 experiment. Is that due to the choice of using RCP8.5 *for initialisation? If so, what effects may this have on the results? If not, what is it due to?*

This is just a plotting error, of not drawing a line between the year of initialization and the first year of output in this simulation.

12. Fig. 1 Please fix in-panel labels in bottom panel (space between parentheses and "dotted"/"solid").

Thanks, we fixed that.

13. Fig. 2 Top row: the black and blue lines are almost indistinguishable. Please make them thicker, use different line styles, or otherwise modify them to make the figure clearer.

The blue and black lines have been made bigger.

14. Fig. 3 Please move the legend to the top panel.

We moved the legend to the top panel, and kept the lifetime information in the bottom panel.

15. Fig. 3 The title of the top panel is chopped off in the PDF I downloaded.

Fixed

16. Fig. 3 In the legend, please use full name of the experiments as done in other figures.

Fixed

17. Fig. 6 Caption: "scenario's" –> "scenarios".

Thanks, we fixed the above.

18. Competing interests: "There is are competing interests at present". Barring the "is are", shouldn't these be stated?

We fix the typo meaning "There are no competing interests at present"

Updated Figures:

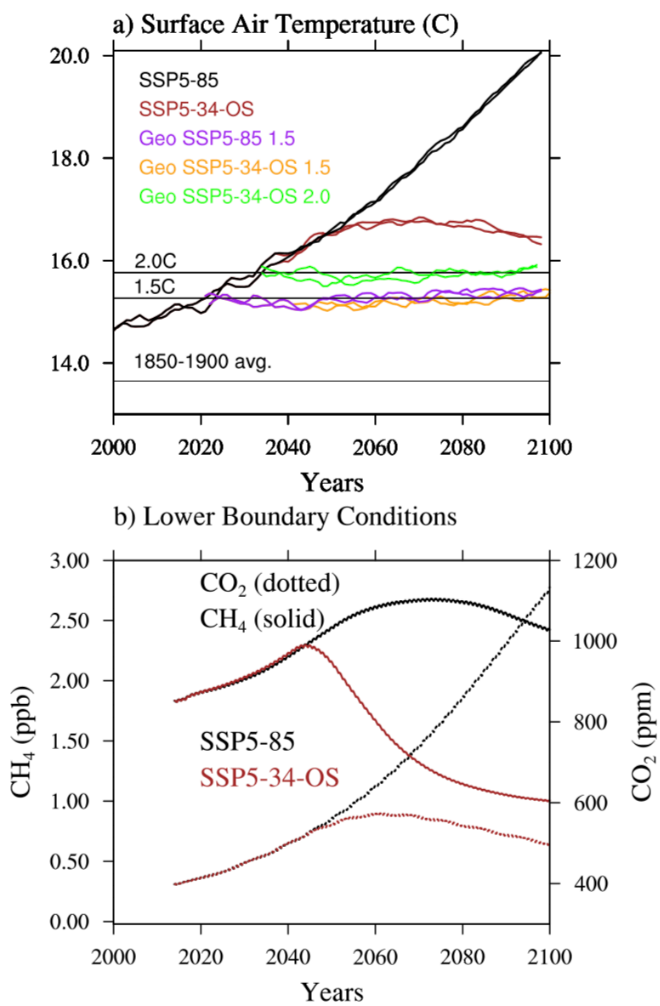

Figure 1. Top panel: Annual surface air temperature evolution for 2 ensemble members of the business as usual case (SSP5-85), the overshoot case that is following business as usual until 2040 and then starting strong mitigation and carbon dioxide removal (SSP5-34-OS), and for 3 different SAG scenarios: based on the SSP5-85 baseline scenario and applying sulfur injections to reduce warming to 1.5^{o}C above pre-industrial (PI) conditions (Geo SSP5-85 1.5); based on the SSP5-34-OS and reducing warming to 1.5^{o}C above PI (Geo SSP5-34-OS 1.5), and based on the SSP5-34-OS and reducing warming to 2.0^{o}C above PI (Geo SSP5-34-OS 2.0) A ten year running mean has been applied to all the timeseries. Black lines indicate the 1850-1900 temperature average (pre-industrial (PI) control temperatures) and the 1.5^{o}C and 2.0^{o}C surface air temperatures above PI control. Bottom panel: Concentrations of carbon dioxide (CO_2), dotted lines, and methan (CH_4), solid line, for the 2 baseline simulations.

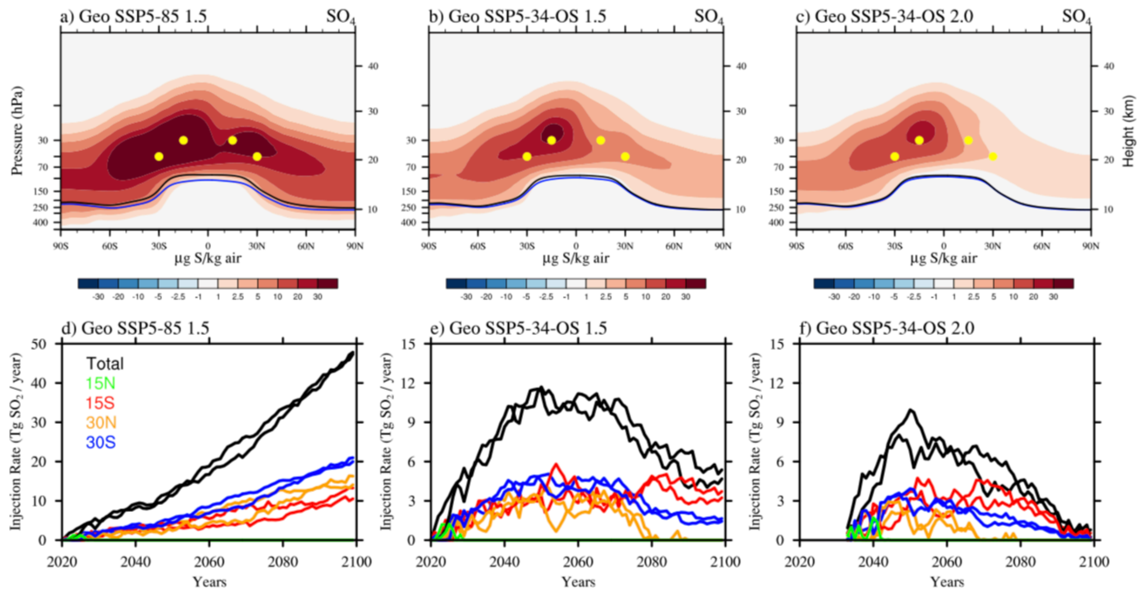

Figure 2. Top row: Difference of zonally and annually averaged sulfate SO_4 burden between the ensemble average of stratospheric sulfur injection cases in 2070–2089 and the control experiment for the same period for Geo SSP5-85 (left), Geo SSP5-34-OS 1.5 (middle), and Geo SSP5-35-OS 2.0 (right). The lapse rate tropopause is indicated as a black line for the control and a blue line for the SO_2 injection cases. Yellow dots indicate locations of injection. Bottom row: Injection rate in Tg SO_2 per year for the three cases as in the top row (including two ensemble members): total injections (black), injections at 15^oN (green), 15^oS (red), 30^oN (orange), and 30^oS (blue).

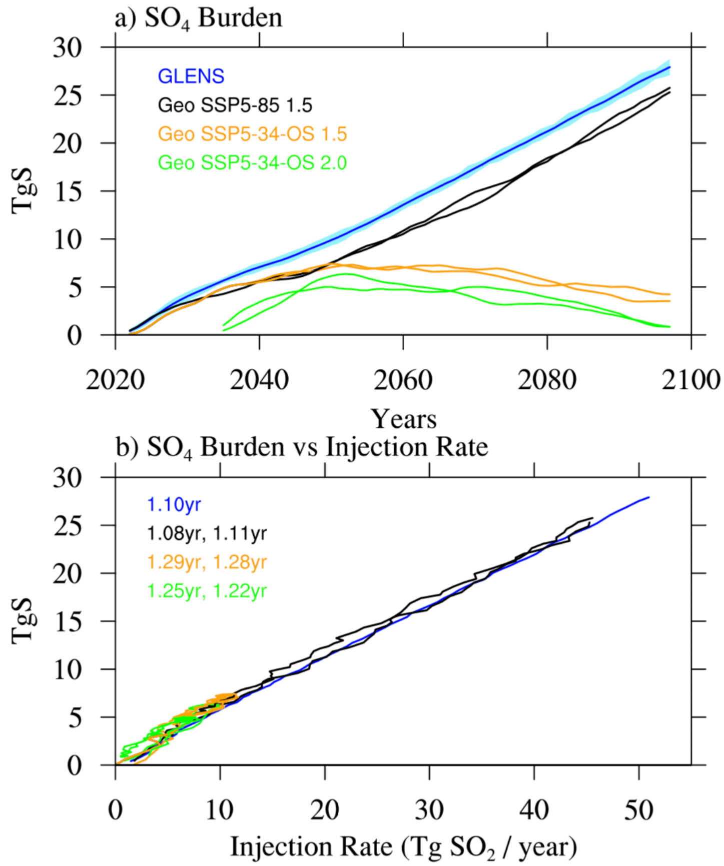

Figure 3. Annual averaged stratospheric sulfate aerosol burden in TgS for the geoengineering injection experiments minus the control with time (top panel) and injection rate (bottom panel). The stratospheric sulfate lifetime is listed in the bottom panel for the two ensemble members of each experiment. In addition to the model experiments performed in this study, we add result for the Geoengineering Large Ensemble (GLENS). See text more more details.

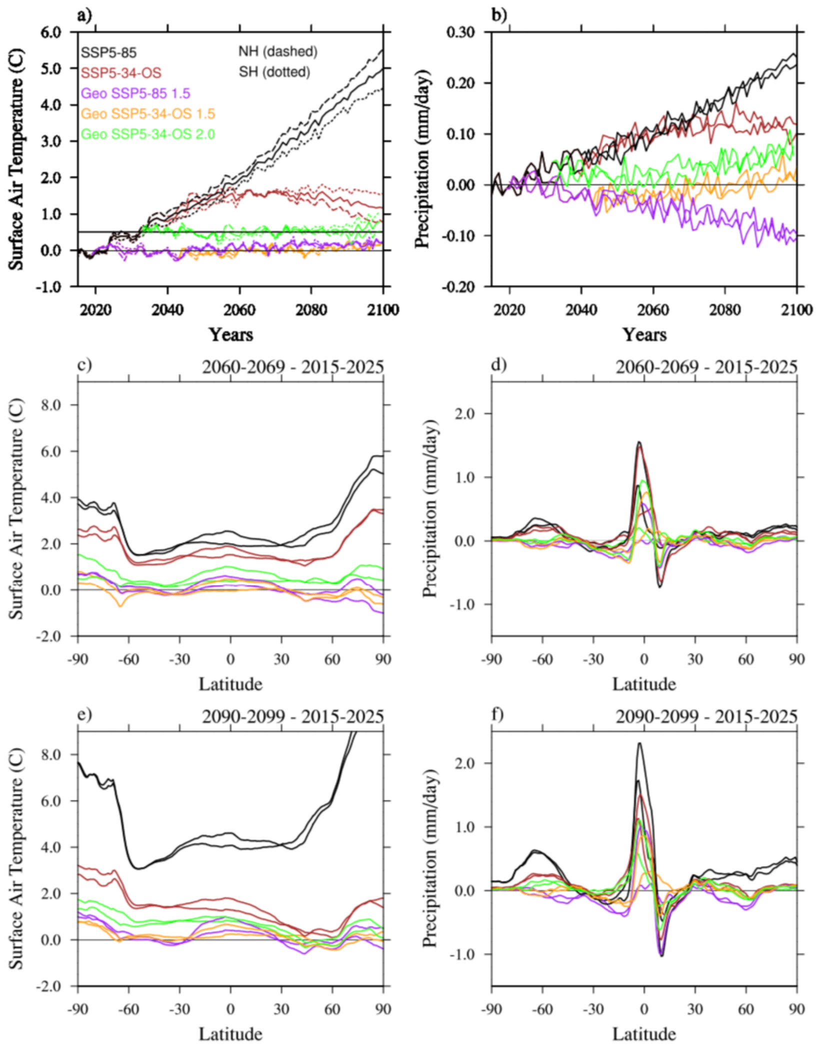

Figure 4. Top left: Time evolution of the ensemble mean area weighted annual mean surface air temperature with regard to 1850-1900 conditions, over the globe (solid), over the Northern Hemisphere (dashed) and over the Southern Hemisphere (dotted) for different model experiments (different colors, see legend); Top right: Time evolution of area weighted annnual precipitation with regard to 1850-1900 conditions for different model experiments and ensemble members (different colors); Middle and bottom row: differences for zonal mean surface air temperatures (left) and precipitation (right) between values in 2060-2069 (middle) and 2090-99 (bottom) for the different model experiments (different colors) and 2015-2025 SSP5-85 conditions.

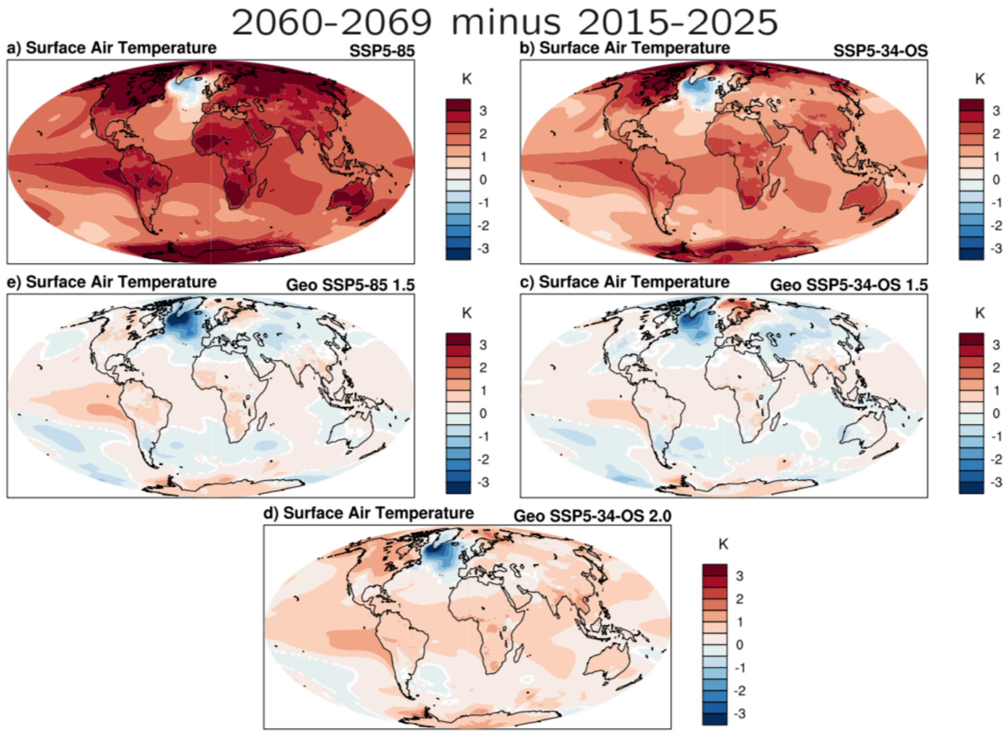

Figure 5. Ensemble-mean surface air temperature difference between 2060–69 and 2015–2025 for SSP5-85 and SSP5-35-OS (top panels), Geo SSP5-85 1.5 and Geo SSP5-34-OS 1.5 (middle panel) and Geo SSP5-35-OS 2.0 (bottom panel).

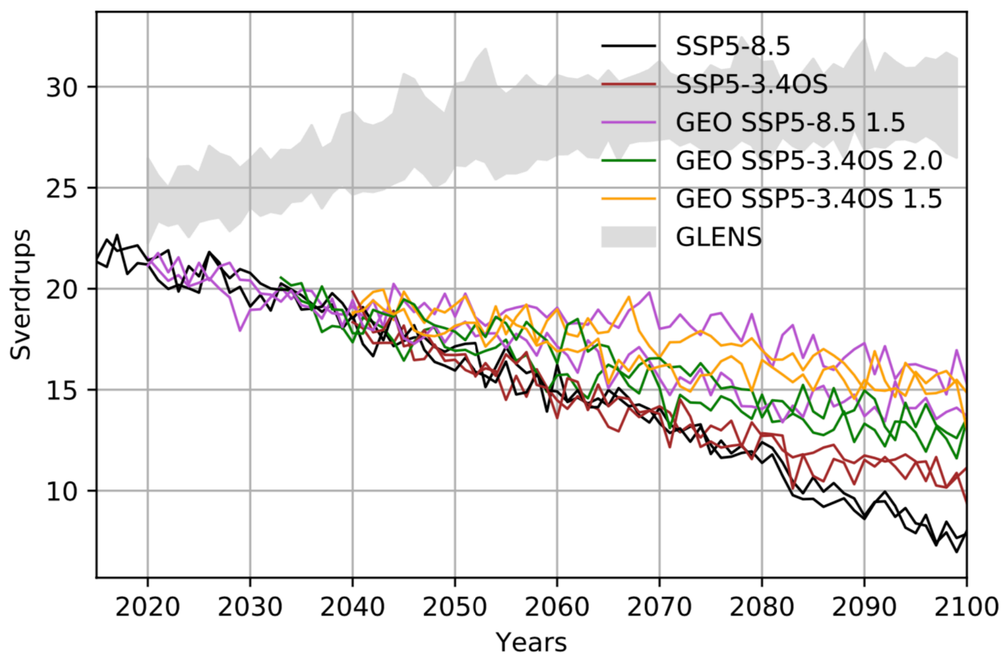

Figure 6. Evolution of the maximum North Atlantic Meridional Overturning Circulation strength from the AMOC index for the different scenarios and ensemble members. Shaded grey area is AMOC index range in the 21-member GLENS ensemble. The AMOC index is defined as the maximum flux in the Atlantic Basin between 500m depth to the bottom, and between 28–90^{o}N (Sverdrups).

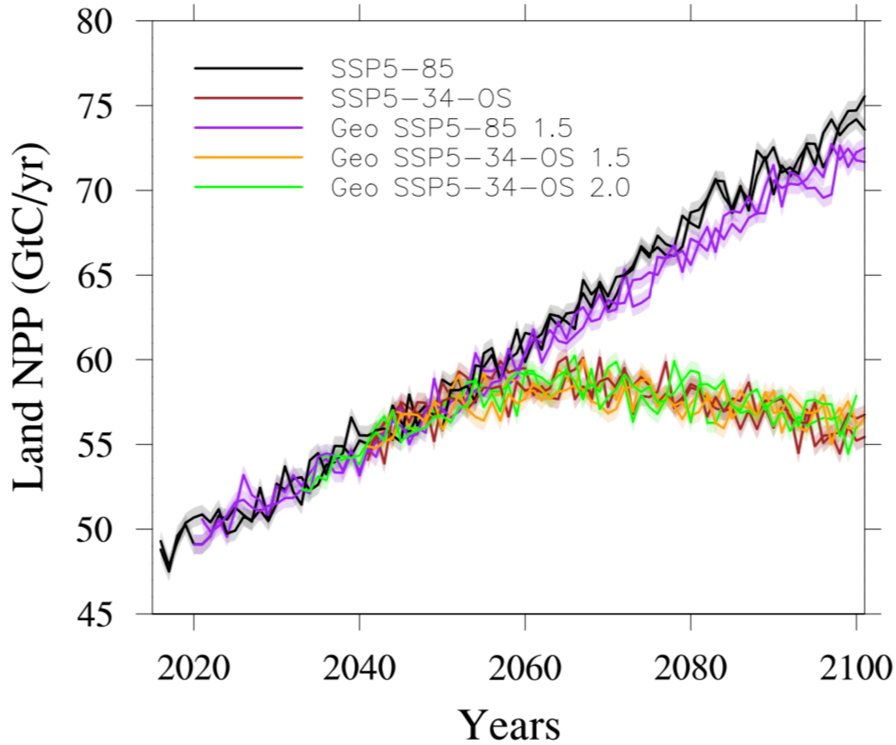

Figure 7. Annual land accumulated NPP (GtC/yr) in baseline and SAG scenarios and ensemblem members (different colors are indicated in legend). The shaded area is 1 standard deviation of 450 years pre-industrial control simulation.

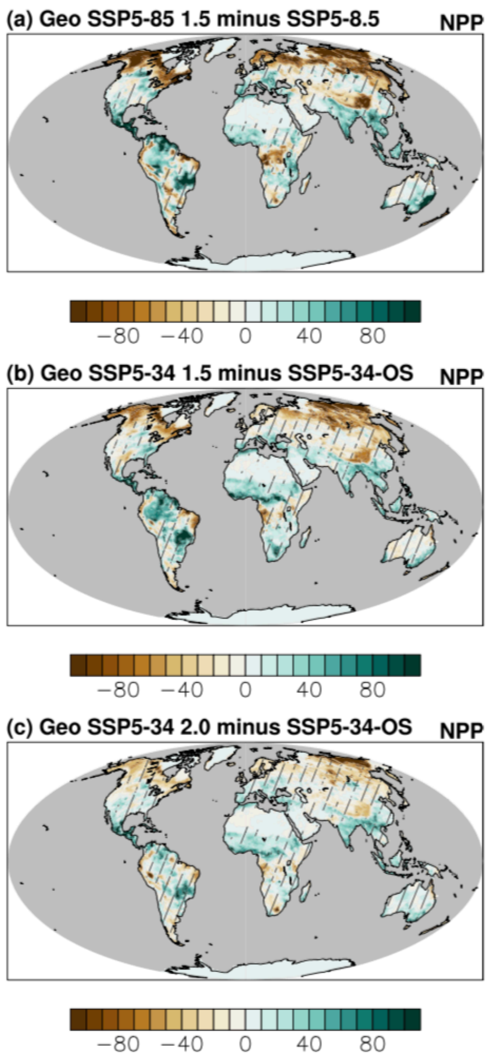

Figure 8. Ensemble mean land accumulated NPP difference (gC/m^2/yr) between 2060-69 for Geo SSP5-85 1.5 and SSP5-85 (top pannel), Geo SSP5-34-OS 1.5 and SSP5-34-OS (middle pannel), and Geo SSP5-34-OS 2.0 and SSP5-34-OS (bottom pannel). Hatched regions are areas with changes within 1 standard deviation of 450 years pre-industrial control simulation.

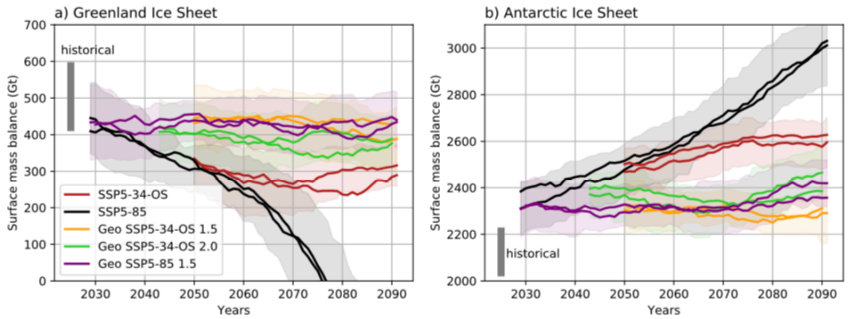

Figure 10. Mean ice sheet surface mass balance (SMB) in Gt per year with shading indicating the standard deviation. A 20-year running mean has been applied to filter out year-to-year variability. For the GrIS (left panel), the area of integration is the contiguous ice sheet (1,699,077 km^2). For the AIS (right panel), the area of integration is the grounded ice sheet (12,028,595 km^2). The solid grey bar indicates the +/- 1 standard deviation SMB over the period 1960-1999 in CESM2(WACCM) for reference.

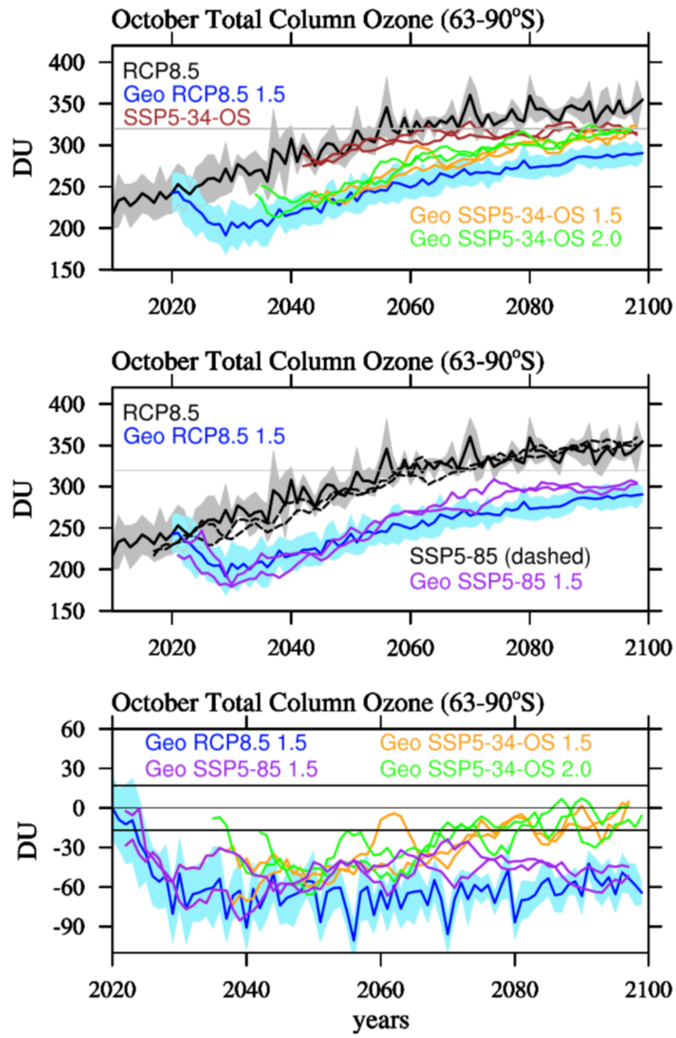

Figure 11. Top and middle panel: October averaged total column ozone between 63-90^{o}S for different model experiments and ensemble members (different colors). Grey and and light blue areas show the standard deviation of the GLENS ensemble and the light grey line indicats 1980 values. Bottom panel, differences between geoengineering and control experiments, the two black lines around zero indicate the standard deviation from the GLENS baseline simulations. A running mean over 5 years has been applied to results from the one-member simulations.