Reviewer #1

The author studied the climate response to carbon capture and stratospheric sulfur injection geoengineering, using simulation results from two coupled global climate models. The author focused on the climate response, including temperature, precipitation, and aridity over major land area especially North and Southern America.

While the scientific question examined here could further our understandings of the potential impact of different geoengineering strategies on the regional climate response, this paper fails to convince me to be useful in a number of aspects (Discussed below).

Therefore, I don't think the current paper is suitable for publishing unless a substantially improved version is made.

Response:

Thanks for constructive feedback on making the paper more useful to the research community. We have made substantial improvements in this revision.

First of all, the way the paper is written could be improved substantially. Many important details that are crucial for readers to understand their results are missing, which confuse me a lot during my reading of the paper. For example, in the Methods section, the model and experiment descriptions are just too simple.

Response:
Points are well taken. We have expanded the model description in the Method section (section 2.1). Even though these two models share a lot of common modules, we now provide more details about them (line 115).

The author compares simulation results from two coupled global climate models, CESM-WACCM and CESM-CAM5, but how different these two models are, and how these differences could affect simulated climate response is rarely included.

Response:
In the previous version, we had mentioned the main difference between these two model versions is the stratospheric component. CESM1-WACCM has a much higher model top and more detailed stratospheric aerosol chemistry and thus is well suited for stratospheric injection simulation. CESM1-CAM5, on the other hand, is the workhorse version of the CMIP5 model and has been used widely (Kay et al., 2017).

We now explained that the model difference can affect the simulated climate response because of the climate sensitivity difference, and we have fully acknowledged this caveat and made great attempts to address it by normalizing the response with respect to the temperature response and emphasizing the underlying physical mechanisms instead of the absolute value of responses.

Also, the author mentioned that the land component differs from each other for these two models, but did not include any discussion about what the difference is.

Response:
The land components are also different between these two versions due to incremental model development. CCR experiment uses CLM4 and GLENS uses CLM4.5. But we do not believe that affects the analysis to a great extent.
The difference between these two models is mainly related to climate sensitivity difference that arises from the cloud parameter tuning during the model development process.

The description of the experiment design and post-processes is also not satisfied. For experiments used in this study, did you use the GLENS project results for the Sulfur Injection case?

Response:
Yes. We did not spell out GLENS in the submitted version but cited the BAMS article promoting it (Tilmes et al., 2018).
This is now corrected.

Where did you get the Carbon Capture simulation results and what is the trajectory of the CO2 concentration/emission in this simulation?

Response:
The carbon capture simulation was done by the first author (Xu) in an earlier study (Sanderson et al., 2017). The trajectory of CO2 emission was explained in Section 2.2b.

In this revision, we plotted the emission cut trajectory and the corresponding radiative forcing drops. The absolute values of CO2 emission and concentration are plotted in Sanderson et al., (2017) so we did not repeat it here.

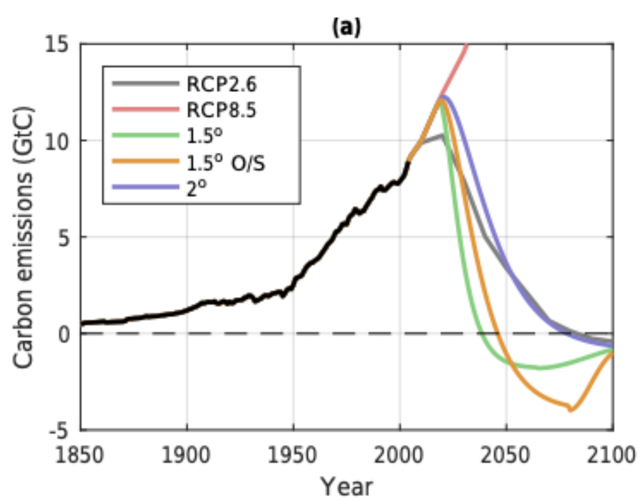

Fig 1a of Sanderson et al., (2017).

For the bias correction, I am confused by how exactly the bias correction is applied to the future simulation cases to account for different climate responses for these two models? etc.

Response:
Bias correction is explained in detail in Section 2.5.
Since the bias of present-day simulation is different for these two models, our bias correction scheme will account for that and apply *different* correction factors to the future projection made by these two different models.

Another example is that in section 2.2, the authors mentioned "CO2 capture and massive CO2 emissions cuts applied to start in 2015-2020", in the next section the author says "both schemes are designed to be deployed in 2020", I am confused which one is accurate? The author should at least state such basic information clearly in the paper.

Response:
Sorry for the inaccuracy. The former is correct.
We have now corrected the latter in Section 3.1 to be "Even though carbon capture scheme is introduced slightly earlier in 2015 to 2020, the longer lifetime of CO2 makes the impact of perturbation to the carbon cycle and atmospheric concentration slower to emerge."

I am also not convinced by the paper. In many places, the author compared the difference of climate responses to two geoengineering scenarios. However, these scenarios are simulated using two different global climate models, and the authors have mentioned that climate sensitivity differs substantially for these two models. How should I believe the difference in climate responses is due to different impacts on these geoengineering schemes, not just because of the difference in the structure of these models? Section 5 is helpful in some way, but I am still not sure how should I interpret most of these results?

Response:
This is the major comment and a deep question.

The different impacts of these two geoengineering schemes are what we set out to quantify. Therefore, we must address the limitation of the current experiment set up – two large ensembles are from two related but different climate models.
As we now increasingly emphasized in this revision, we highlight four approaches to minimize this limitation:
(a) bias correction (see a more technical response below),
(b) normalization (Section 5 as the reviewer acknowledged),
(c) interpretation of physical mechanism especially the role of solar dimming at the ground surface which is only strongly operating in the sulfate injection case (Section 4)
(d) further corroboration of the physical mechanisms at play, using other previously published simulations including volcanic eruption and tropospheric aerosols (Section 3).

After all, we call for more coordinated experiments to systematically examine various geoengineering schemes (including carbon capture), which is currently missing in the literature.

Many descriptions presented in the paper seem not accurate to me.

Response:
Most of the minor comments below involve inaccurate statements in the original submission. We now made clarification as recommended by the reviewers.
Thank you!

For example, the authors claim that "spatial patterns show similar agreement between P/PET and soil moisture as well as P-ET" in figure 2. While the time series do show similarities between different indexes, the spatial patterns show opposite responses in many regions, especially Eastern North America and Southern South America. I am not sure if this could be seen as similar?

Response:
Yes. The purpose of Fig 1 and 2 is to justify the use of P/PET as the main metric since we do not want to present an analysis that's too complex using multiple drought metrics.
The stronger similarity in time series in Fig 1 is expected because the numbers are the average over a relatively larger region.
In contrast, the detailed map presented in Fig 2 exposes fine details of these metrics and their subtle differences over smaller regions, as expected (e.g. Eastern North America and Southern South America as the reviewer correctly pointed out).
However, we stress the overall similarity between responses of P/PET, soil water, and P-ET at continental scale including western US and Amazon regions (boxes in Fig 5a) where the drying trend is largest and passing the significant test.

Line 7, the authors claim that "it is essential to apply the bias correction to model output before carrying out a meaningful comparison of future changes". If you are using simulation results from the CESM_RCP8.5 2010-2019 period to correct both WACCM_RCP8.5 and WACCM_SulfurInjection, and since WACCM_SulfurInjection is the same to WACCM_RCP8.5 before 2020, why should the difference between these two cases change compared to results without bias correction?

Response:
The difference between WACCM_RCP8.5 and WACCM_SulfurInjection will yield the "impact" of sulfate injection. The result will change compared to raw data without bias correction because, for precipitation and PET, the bias correction is done via multiplication/division, not simple addition/subtraction.
In the case of T itself, the reviewer is correct that the bias correction applied to future simulation will not change the difference between the pair and will yield identical results regard to the impact of sulfur injection.
We now make the technical note above in section 3.1.

Can you compare these simulations using the same model for all these simulations? Otherwise, I am not sure what is the best way to account for difference caused by using different models.

Response:
That will be the dream simulation we are proposing in the end!
Please see the response to the major comment above.

Line 10-12: First of all, I think the author should emphasize more that the analysis here is only for changes over land, especially when you are presenting the results.

Response:
Yes. Although we spelled out "major land region" in the figure legend, we did not explicitly state it in the text. This is now corrected.

The author claims that the larger warming over land for WACCM is due to the larger climate sensitivity of the model. The two models differ not only in climate sensitivity but also the ratio of the land to ocean temperature responses. How much of this larger warming is due to larger climate sensitivity, and how much is due to different land/ocean temperature responses?

Response:
This is a good point and we checked as suggested. The land/ocean ratio for WACCM and CESM are both ~1.5, thus we continue to emphasize the main discrepancy coming from climate sensitivity due to model structural differences, and we continue to try to mitigate this caveat by focusing the discussion on the relative values and physical mechanism, as opposed to absolute values which are model-dependent anyway and thus less useful when broadening this kind of analysis to multiple models.

The same concern applies to the next paragraph (lines 16 to 20). How much of the larger warming in WACCM compared to other models is due to climate sensitivity or land/ocean responses in these models? If the main reason is climate sensitivity, then the larger cooling in WACCM could just because WACCM produces more warming and the goal is to stabilize global mean temperature changes at the same level. I don't think the reason is simply that a larger injection amount or the injection strategy. If the main reason is the different land responses, then does the explanation in the previous paragraph using climate sensitivity still hold?

Response:
It's hard for us to trace down other sulfur injection model experiments to verify the land-ocean warming ratio, so we have removed the statement of WACCM cooling being larger than others. But we do note that the sulfate injection here is introduced off tropics so that a larger cooling is expected while keeping the injection amount the same. The revised paragraph reads:
"The Sulfur Injection simulation here leads to a cooling of 6ºC towards the end of the century, compared with the baseline warming. This larger cooling is designed to largely balance the projected warming by introducing a large amount of sulfur gas. Due to the experimental design that the injection location is off the tropics, this model would produce a greater temperature response than the injection solely from the tropics using the same model (Tilmes et al., 2017; and Kravitz et al., 2018)."

Also, I don't know what you mean by saying "the longer lifetime of CO2 makes the impact of perturbation to the carbon cycle and atmospheric concentration slower to emerge". Factors that affect the cooling effect of CO2 capture include the inertia of the ocean, the response of the global carbon cycle, etc. I don't know what does the longer Co2 lifetime means here.

Response:
This is exactly the type of physical mechanism distinction we want to highlight between two types of geoengineering approaches. So thanks for asking.
It's well known that compared to the duration of aerosols floating in the stratosphere from months to years, the lifetime for CO2 is about decades to centuries. What that means to the radiative forcing perturbation (which is proportional to CO2 atmospheric concentration) is that its response to the emission change (or capture) is much slower than aerosol's response. This is illustrated in Figure 4 (e and f).
The inertia of the ocean applies to both cases, and that's why we said "*additional* inertia in CO2 mitigation…)

The last sentence point to Table 2, column 3. I don't know where to find these results?

Response:
Sorry. This statement refers to an earlier table in which global temperature and land temperature are reported separately. Now we have deleted this statement.

Also, the author saying "if only the land variables are considered". Do you have any explanation for this?

Response:
The cooling as a fraction of the baseline warming is larger over land compared with the global average, is due to the land-ocean warming contrast ratio as we discussed above. This is not particularly sensitive to forcing introduced (carbon capture or sulfur injection), and thus we chose not to continue to emphasize it.

Line 54 to 56: From the time series plot, you showed a decrease in precipitation by comparing results between WACCM_SulfurInjection (2086-2095) - WACCM_SulfurInjection (2010-2019), but in table 2, the decrease in precipitation for end of century is 121.0 mm/day compared to WACCM_RCP8.5, smaller than the increase in WACCM_RCP8.5 compared to present day (123.2 mm/day) value. If I understanding this correctly, the present-day value should be the same for WACCM_RCP8.5 and WACCM_SulfurInjection because they branched from each other in 2020. So why is there a decrease instead of a slight increase in precipitation (am I reading this wrong here)?

Response:
Thanks for catching the details. Yes. We meant that looking at the blue line in Fig 3d, WACCM_SulfurInjection precipitation is trending down (slightly) from <u>2020 to 2095</u>. Comparing 2086-2095 with <u>2010-2019</u>, the reviewer is correct in pointing out the WACCM_SulfurInjection will actually see a negligible increase of 2.2 (123.2-121.0)

mm/year, mainly due to the increase of about 10 mm/year during the ten years prior to the injection starting at 2020.
The quantitative statement is now revised at the end of Section 3.2, without impacting the argument.

Line 102-104: It is clear from Fig.4d that the P/PET index has different changing rates before and after 2050 and since changes in PET is quite consistent throughout the simulation, the difference is mainly caused by different responses in P (Maybe you should do the same calculation and see how much these two terms contribute to the difference in P/PET).

Response:
Yes. We have now calculated the changing rate of P and PET separately to aid the interpretation of P/PET changes.
Please see the revised Fig 4 and the numbers quoted in Section 3.3 (the last paragraph).

However, the authors attribute these differences in P simply to "when the surface cooling starts to emerge, the precipitation also weakens in response to Sulfur Injection". What do you mean by saying "the surface cooling starts to emerge"? The changes in P/PET results seem interesting to me, but I don't think the authors explain these changes in an appropriate way here. I don't even understand what you mean by saying "when the surface cooling starts to emerge" here?

Response:
This refers to the statement at the end of Section 3.3, discussing the response of P/PET in the near term vs. long-term (before and after 2050), and how it differs from Sulfur Injection and Carbon Capture. The phrase "when the surface cooling starts to emerge" is too vague (sorry for the confusion), so we have written the paragraphs completely. It's partially copied here for easy references:
"This non-monotonic behavior in Sulfur Injection induced a P/PET response that is highly distinct from the Carbon Capture case in terms of timing. The latter, in contrast, always falls behind the Sulfur Injection changes in inducing climate responses (green curve vs. blue in Figure 4a-d). The P/PET enhancement due to Carbon Capture (compared relative to baseline warming, not relative to present-day), only starts to be significant toward the end of the century with a growth rate of 1.4%/decades after 2050, almost three times larger than the decades before 2050."

Section 5 investigated the contribution of P and PET to the change in P/PET for different scenarios. For the discussion of P, you compared the precipitation results of stratospheric sulfur injection with tropospheric SO_2 increase, and then say "the larger slope of precipitation for Sulfur Injection is expected." Could you directly compare these two results under different forcing designs? I am not sure how could you then get the conclusion that "the model dependence is the largest uncertainty" when you are not comparing the same kind of forcings?

Response:
Thanks for catching that. The discussions related to the synthetic analysis combining the current data with our previously published results (using 20th-century simulation).

We have now removed the statement on "the model dependence is the largest uncertainty" because of a lack of justification.

Also, for the regression results, did you check the standard deviation for these regression results? How robust the regression results are?

Response:
It's unclear whether the reviewer is referring to Table 4 or Figure 8. In Figure 8, the uncertainty range of regression lines is reported in the figure header, with the PET regression onto T being the tightest one. In Table 4, similar statistics had been shown in Lin et al., (2016) so we did not repeat them here.

It is interesting that global major land PET scales almost linearly with global mean T. If the PET is such tightly related to the T change, why is there a difference in PET slope between these two forcings? I don't think the authors explain it in a reasonable way. If you attribute this also to model difference, then maybe you should use the same model to do the comparison.

Response:
This seems to refer to Fig 8b. Yes, you are right, global major land PET scaled linearly with global mean T. But still, different forcing (SO2 or CO2) can leads to different PET sensitivity. Lin et al. 2015, JGR has shown using the same model, GHG and anthropogenic aerosols, CO2, black carbon, SO2, can all impact PET through different mechanisms, notably the role of surface available radiation. See below.

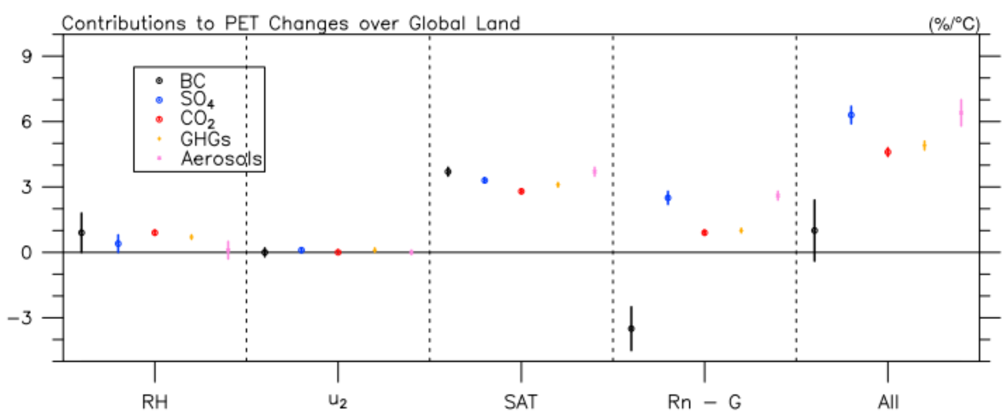

Figure 3. The effects (in percent change of PET per degree) of relative humidity (RH), wind speed (u_2), surface air temperature (SAT), and available energy (Rn-G) on the percentage changes in potential evapotranspiration (PET) over global land due to black carbon aerosol, sulfate aerosol, and CO_2, scaled by global mean surface air temperature change. The results of greenhouse gases (GHGs) and aerosols from L. Lin et al. (submitted manuscript, 2015) are shown for comparison. The error bar denotes two standard deviation.

This decomposition of PET changes is now repeated in Table 5 and Figure 9 to better explain the different slopes of PET in a reasonable way.