Referee #2

Xu et al. compare the changes in aridity under a worst-case warming scenario (RCP8.5), an extreme mitigation and carbon dioxide removal scenario (that I'll refer to as CDR) and the GLENS sulphate aerosol injection scenario (that I'll refer to as GLENS). They combine results from 2 versions of the same base model (with importantly different climatologies) CESM1, which ran the CDR scenario, and WACCM, which ran the GLENS scenario, to assess the surface hydrological response to these different scenarios. To make the two ensembles comparable bias-correction is applied such that both models are adjusted to match the observed climatology.

The authors evaluate a range of hydrological measures and settle on P/PET as their choice of aridity measure. Focusing on this metric they evaluate changes over the global land area (I'm guessing this is what is referred to as "major land") and across the Americas, and evaluate the differences between the CDR and GLENS response. They also normalize the response by global temperature to address a mismatch between the CDR and GLENS experimental objectives.

They find that RCP8.5 produces a general aridification over the land and over the sub-regions they focus upon, driven by a large increase in PET with regionally divergent P responses. They find that while CDR generally partially offsets these trends, GLENS goes too far producing a substantial net reduction in PET that more than offsets the reductions in precipitation. The Amazon is an exception for GLENS, where instead of producing a net increase in P/PET instead it only partially offsets the substantial reduction seen in RCP8.5.

Response:
Thanks for the excellent summary which says it all.

General comments

My main reservation about this paper is that I don't believe comparing these scenarios is policy-relevant or that scientifically useful. The title sets this up as a competition (vs.) between two geoengineering options but these policies are complementary.

Response:
This is a good point. It's not our intention to set up this comparative study as a competition and selection practice. We have changed the title to reflect this " a comparison between Carbon Capture and Sulfur Injection"

The CDR scenario (carbon capture as it's often referred to) is also not primarily a CDR scenario but largely an emissions reduction scenario as compared to RCP 8.5 with net negative emissions only after 2050.

Response:
Even though a net negative emission is achieved after 2050, the CDR capacity is assumed to scale up between 2020 and 2050, gradually taking over the contribution by emission cut. For the purpose of comparing the climate response, we did not aim to differentiate the extract contribution of both.

We now acknowledge the definition in Section 2.2.b *(We show the mass amount of CO2 reduced (either through emission cut or various ways of carbon capture) and the associated negative radiative forcing in later figures to illustrate the impact of these idealized trajectories."*

The other issue is that the experiments were designed to achieve different objectives and produce quite different radiative forcing and temperature responses over the analyzed period and were run in model-versions that differ in quite important ways. Both of these factors undermine the utility of this comparison.

Response:
We acknowledge that the two model versions are not exactly the same, which may hinder some analysis. We had made attempts to mitigate that as detailed in later responses. It is our ultimate goal to inspire more coordinated model experiments (from multiple groups hopefully) in the future, which is currently missing.

Despite the lack of coordination when these two sets of the simulation were produced a few years ago, we do not agree that these two experiments were designed to achieve fundamentally different objectives. They both aimed to achieve a temperature stabilization towards the end of the 21st century (at slightly different levels per Fig 3c), and this long-term equilibrium provides a great opportunity to study climate response.

That said, clearly a lot of careful analysis has been done and so I'll continue to my other comments.

Response:
Thanks for acknowledging the comprehensiveness of our analysis.

Much of the paper is written as if the two experiments had the same aim but they didn't. The methodology doesn't make the different goals of the two experiments clear enough and their high-level difference ought to be described up front, i.e. section 2.2 should be elaborated upon to make clearer the experimental setups and their differences.

Response:
The suggestion of making the experimental objectives clear is well taken. We now acknowledge that the two experiments were designed and conducted by different groups upfront in the new opening paragraph of Section 2.2.
Two pairs of model simulations were conducted and published in the last few years, both featuring a large ensemble to enhance the robustness of examined climate responses, especially at a regional level. The WACCM set (http://www.cesm.ucar.edu/projects/community-projects/GLENS/) is aimed at stabilizing the climate at its 2020 level with sulfur injection. The CESM set (low warming large ensemble, http://www.cesm.ucar.edu/experiments/1.5-2.0-targets.html) is aimed at climate stabilization at 1.5 and 2ºC warming levels, with aggressive emission cut and rapid growth of negative emission technology.

There is often reference to “lagging effects” of the CDR experiment but given that it had different ends from those of the GLENS experiments it seems inappropriate to describe the temporal evolution as if it were trying to achieve the same thing.

Response:
The “lagging effects” of CDR is a point we want to emphasize so thanks for bringing it up. As shown in Fig 4 (the temporal evolution of perturbation and response), the CDR measures actually lead the Sulfur Injection (Fig 4e), so there is indeed a lagging in the achieved forcing (Fig 4f). Note that the temperature response (Fig 4a) to the forcing will also lag (due to ocean thermal inertia), but irrelevant to the forcing imposed.

I was surprised that transpiration wasn’t discussed as this is a major driver of terrestrial hydrology. I strongly recommend considering the direct CO2 physiological effect of CO2 in these simulations as this will be a big difference between CDR and RCP8.5 / GLENS.

Response:
Good point, and we add a note in the Conclusion section.
The direct physiological role of CO2 is potentially important because the CO2 level is greatly reduced in only one of the two mitigation approaches. But this study, focusing on meteorological drivers of land aridity using P/PET, did not delve into CO2’s suppression on plant transpiration via stomatal closure, which also appears to be weak in these two models compared with other climate models (Swann et al., 2018)
The weaker physiological effect is in response to comment later.

As for the calculation of PET, we formulate it as a function of temperature, relative humidity, available energy, wind, <u>and bulk stomatal resistance (rs)</u>. We indeed employed a fixed bulk stomatal resistance (i.e., 70 s/m) while those values can vary over different regions and time as the reviewer suggested.
The simplification is justified by that even adopting widely different but reasonable choices of vegetation parameters including rs, the PET values do not change much (within 5%). Scheff and Frierson (2014), Fu and Feng (2014) and Fu et al., (2016) showed:

The change of PET by neglecting the higher-order terms can be written in the form

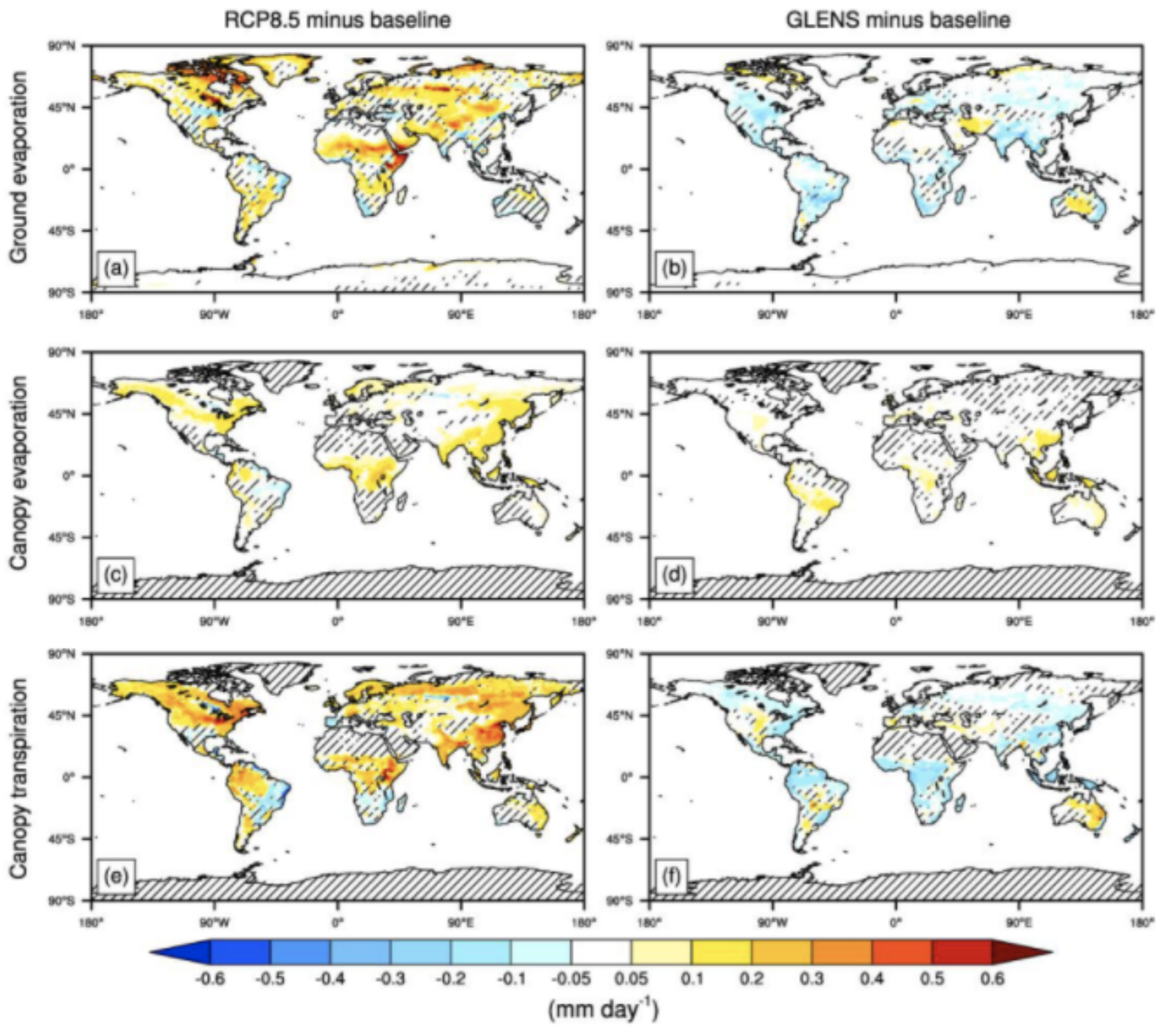
$$\Delta \mathrm{PET} \approx \frac{\partial \mathrm{PET}}{\partial \mathrm{SAT}} \Delta \mathrm{SAT} + \frac{\partial \mathrm{PET}}{\partial \mathrm{RH}} \Delta \mathrm{RH} + \frac{\partial \mathrm{PET}}{\partial (Rn - G)} \Delta (Rn - G) + \frac{\partial \mathrm{PET}}{\partial u2} \Delta u2 + \frac{\partial \mathrm{PET}}{\partial r_s} \Delta r_s \qquad (1)$$

$$\frac{1}{\mathrm{PET}} \frac{\partial \mathrm{PET}}{\partial r_s} = - \frac{\gamma C_H |u2|}{[\Delta(\mathrm{SAT}) + \gamma(1 + r_s C_H |u2|)]}, \qquad (2)$$

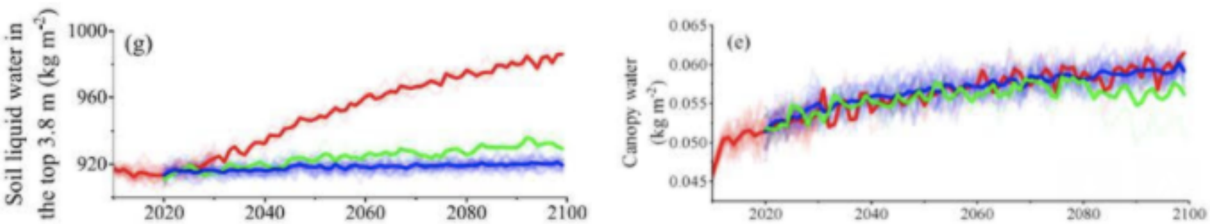
Note that the magnitude of the PET relative change is largest when rs=0, then it goes down to be less than -0.4%/ (s/m). The negative sign means that, with elevated CO2 level, a larger stomatal resistance leads to smaller PET (implying a terrestrial wetting if everything else is fixed).

The authors report results showing PET and ET over land, finding that they both increase in lockstep. This was surprising to me, given the results of Swann et al. (2018) who found that PET was projected to increase in models while ET was not. Swann et al. found that this was due to reductions in transpiration due to CO2 fertilization offsetting the meteorological drivers that drove PET increases. This suggests that the CO2 physiological effect has not been included or is very weak. Please provide details and analysis to address this.

Response:
This refers to Fig 1(a). ET increase in WACCM simulation but note that ET increase is 100 mm/year less than increase in PET). The ET increase is also consistent with decrease in soil water in top 10 cm of soil.
But, Why an increase in ET? Cheng et al. (2019), co-author of the present study, has looked into that carefully and found that global mean ET is indeed increasing including the canopy transpiration part under RCP8.5. We also checked distribution of canopy transpiration (Bottom row in the next figure), with increase in most land area under RCP8.5, except Southeast Brazil and Southern Africa.



As speculated by the reviewer, in these two models, increases in soil liquid water and canopy intercepted water under RC8.5 (red lines in the next figure) overwhelms the direct physiological effect of CO_2 (suppressing transpiration) in RCP8.5.



The different model behavior is also one reason that we did not delve deep into analyzing other direct land model output.

Relatedly, it is not clear whether there were any differences in land cover between the CDR scenario and the RCP8.5 / GLENS scenarios. This could have a large effect on regional hydrology.

Response:
To our knowledge, land cover and land use are not perturbed in CDR or GLENS scenarios and thus should be consistently the same as in RCP8.5. We add this note to Section 2.2.

The temperature difference between the two experiments should be made clear, from table 2 I can back out that CESM-1.5C is 0.7C / 0.5C warmer than the baseline whereas GLENS is 0.1C warmer (is this right? I thought it was designed to perfectly offset warming from the baseline period).

Response:
Those are correct. GLENS is designed to be stabilized at the 2020 climate, while our baseline is 2010-2019 and thus the 0.1C difference.

Given the lower sensitivity of CESM, this 0.5C figure would be larger if both were run in WACCM. This could be driving some of the difference between these two experiments and should be brought out and discussed.

Response:
Not necessarily. The 0.5ºC is the experiment design that aims to stabilize the climate at 1.5ºC inspired by the Paris Agreement. (the 2015 warming is 1ºC relative to pre-industrial) Had the WACCM been used for this simulation, we would change the emission trajectory to meet the 1.5C level as well so to maintain the 0.5ºC difference.

I'd like a clearer sense of the magnitude of the bias-correction overall and of the regional character of it in your study region. The fact that WACCM has around 25% more precipitation over "major land" (undefined as far as I can tell) ought to be highlighted!

Response:
Let us clarify.

The magnitude of the bias-correction overall can be seen from Fig 3 (first two rows). The magnitude of the bias-correction of the focused regions can be seen by contrasting Fig 2 (uncorrected) and Fig 5 (corrected).
From Fig 2b, WACCM has around 5% more precipitation than CESM, not 25%.
Te "major land" was defined in Table 2: "the land regions over 60ºS to 60ºN, thus excluding cold regions where the seasonal snow cover or permanent ice sheet surface makes P/PET less useful as a predictor for aridity and vegetation types."

Beyond this the regional biases ought to be made clearer. Are the models far off in key regions such as amazon? How wrong are the arid and semi-arid areas calculated using the models uncorrected P/PET values? How different are the regional biases? An extra figure or two is needed to make this clear.

Response:
Thanks for asking these questions.
Comparing the observation in Fig 5a, and the uncorrected models in Fig 2a and Fig 2g, the models appear to be doing a reasonable job in capturing present-day P/PET, over Amazon, and also over the arid/semi-arid transitional regions. Of course, the corrected model in Fig 5b and c have matched with observed P/PET nicely.
Given the number of tables and figures already included, we hesitate to include more display items.

The temperature normalization section doesn't seem to add much to the paper and has some serious problems. I understand that it could be useful for comparing scenarios with different levels of cooling if it were not for the difference in the climate sensitivity. WACCM's climate sensitivity looks to be ~20% higher than CESM's which will mess with this normalization procedure.

Response:
Section 5 on normalizing the regional change relative to global cooling numbers is specifically added to address the caveat that the two models have different climate sensitivity.
Our rationale is that normalization can be used to tease out both the forcing differences (as in our previous paper of Wang et al., 2017 Sci Report using a single climate model), as well as model differences.

I'd suggest either cutting this section (what do we learn that isn't covered elsewhere?) or addressing this climate sensitivity issue by testing how different the RCP8.5 sensitivities are between these 2 models. I worry that model differences rather than scenario differences could be driving some of the response seen in this section.

Response:
Actually as argued by Reviewer #1, the model difference rather than scenario difference will affect the absolute values of the response reported in previous sections to a greater extent than in this normalization section. Therefore, it is our intent to keep Section 5 as a

mitigation to the model difference issue. We also elaborate on this rationale in the Conclusion section.
"An important note should be made regarding the interpretation of the quantitative "benefits" presented. Because these two models have different climate sensitivities, the baseline warming induced by unchecked emissions growth is not at the same level. The WACCM warms faster, reaches the 2ºC level at earlier decades, and has higher end-of-century warming at 6ºC, compared to the CESM. Since the baseline warming is different for the two models at different decades, the climate "benefit" due to any mitigation measures shall also be interpreted in a relative sense, i.e., the fraction (%) of the projected change in the future that can be mitigated by the Carbon Capture or Sulfur Injection. For example, even if our results suggest Sulfur Injection can lead to 6ºC cooling while the Carbon Capture can lead to a 4ºC cooling, that does not quantitatively provide any constraints onto the strength of respective approaches. "

The section summaries seem unnecessary.

Response:
Because of the lengthy mixture of results and discussion in Section 3 and 4, we still keep the section summary for clarifying the main points.

I think it would be more fair to describe the "carbon capture and storage" scenario as an extreme mitigation scenario. A reduction in positive emissions makes up the bulk of the difference between RCP8.5 and this scenario.

Response:
We agree. We now clearly explained in Section 2.2 that the CESM experiment is aiming to test the impact of an "extreme mitigation scenario" that contains a mixture of aggressive decarbonization and carbon removal.
"The CESM set (low warming large ensemble, http://www.cesm.ucar.edu/experiments/1.5-2.0-targets.html) is aimed at climate stabilization at 1.5 and 2ºC warming levels, with aggressive emission cut and rapid growth of negative emission technology, an extreme mitigation scenario."
We also note that for the physical climate system to respond, one tonne less emission would be the same as one-tonne carbon captured.

If you wish to highlight the use of carbon dioxide removal (CDR) or the presence of negative emissions then I'd suggest using these terms instead of carbon capture and storage as this technology can be used without producing negative emissions, i.e. on coal power plants.

Response:
The negative emission technology has too many different terminologies already these days, carbon and storage, carbon capture and sequestration, carbon dioxide removal, direct air capture. We do want to avoid getting into further complexity and thus retained the use of "carbon capture" as a short-hand.

But, in Section 2.2b, we thoroughly clarify what we really meant by "carbon capture".

"with CO_2 removal and massive CO_2 emissions cuts applied to start in 2015-2020 (hereafter referred to simply as "Carbon Capture")
"We show the mass amount of CO_2 reduced (either through emission cut or various ways of net emission technology) and the associated negative radiative forcing in later figures to illustrate the impact of these idealized trajectories. However, the technical feasibility and required socio-economic shift to facilitate the scale-up of capacities such as direct air capture (Hanna et al., 2020) and clean energy transition (Hanna et al., 2020b) is beyond the discussion of this paper."

Also in the abstract, we now clearly state: "The CO_2-based mitigation simulation is designed to include both emissions cut and carbon capture."

I'd suggest coming up with some clear shorthand for the experiments and using it consistently in both text and figures, e.g. CESM-RCP8.5, CESM-CDR, WACCM-RCP8.5, WACCMGLENS, Baseline.

Response:
Yes. The shorthand is now consistently used as CESM-RCP8.5, WACCM-RCP8.5, CESM-CarbonCapture, WACCM-SulfurInjection, CESM_historial.

The figures have inconsistent labeling, line colours and styles and some of the captions are oddly formatted. The manuscript text needs a careful proof-read by a native English speaker. There were too many grammar mistakes so I only addressed the worst. There were also many very short paragraphs that could be merged with their neighbors.

Response:
The figures and text will be thoroughly revised in the next version. Short paragraphs were merged where appropriate. Thanks for the careful comments below.

Specific Comments

N.B. Specific comments are given in the order that they appear in the manuscript with the line numbers as I saw them. It seems the pdf has cut off the hundreds part of the line number and I haven't converted the cycling line numbers into something more sensible.

Response:
Yes. We use the cycling line numbers in the published pdf.

Title – Given that you stress in the conclusion that you aren't trying to evaluate which is better and instead are focused on the mechanisms, I'd suggest: "climate engineering and aridity in the Americas: a comparison of carbon dioxide removal and sulphate aerosol injection"

Response:
We agree and we change the 2nd part of the title to avoid the impression that the two are in competition. It is now "a comparison between Carbon Capture and Sulfur Injection".
We stick with carbon capture as explained previously. We did not use SAI (a common shorthand though) because technically what's injected is gas, not aerosol.

22 – I'd argue that this paper hasn't used a consistent framework: different models and different objectives for deployment

Response:
We now tune down the language a bit.
It now reads: "Despite being in the public debate for years, these two leading geoengineering schemes have not been examined under a consistent analytical framework using global climate models. "

22-24 - Given that this is more a mitigation scenario than a pure-CDR scenario, it's not correct to describe this as the first paper to compare sulfate geoengineering against "carbon capture" as previous studies have made such comparisons: [Niemeier et al. 2013: DOI:10.1002/2013JD020445; Muri et al. 2018; DOI:10.1175/JCLI-D-17-0620.1; Jones et al. 2018, DOI: 10.1002/2017EF000720, etc.].

https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2013JD020445
https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-17-0620.1
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017EF000720

Response:
Thanks for suggesting the references, which are now cited in the conclusion section to provide a broader context.
" (2) provide broader insights to the mitigation impact of other geoengineering approaches beyond the two discussed here (such as cirrus ice cloud thinning or marine warm cloud brightening, Muri et al., 2018; and surface albedo modifications, Crook et al., 2015; space mirror, Niemeier et al., 2013). "

Of the three papers suggested, only the Jones et al., (2018) is somewhat close to our analysis, in which SRM is indirectly compared with RCP2.6 which has a weak deployment of CDR. Note that CDR in RCP2.6 is much weaker than what is in our CESM experiment (1.5C).

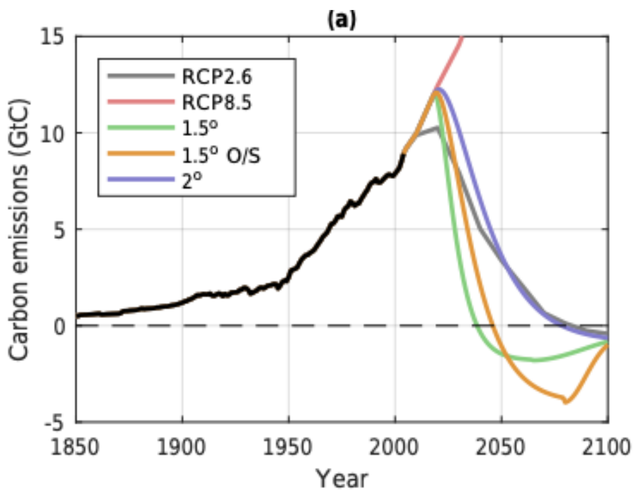


Fig 1a of Sanderson et al., (2017).

But even Jones et al., (2018) did not explicitly compare SRM against the carbon mitigation scenario as suggested by the reviewer.

32 – what does "mitigation potential" mean in this context?

Response:
It means the potential mitigation benefits via climate engineering. Since all of these are hypothetical by design, the benefit or impact is now unrealized potential. We change it to potential mitigation benefits in this revision.

39 – does it worsen the trend or is the trend worsening under RCP8.5?

Response:
The trend is worsening under RCP8.5. The sulfur injection is not very effective in curbing it.

67 – Vaughan and lenton 2011 don't provide evidence of investment, do they?

Response:
We remove the statement of seed investment in this revision.

68 – I thought CDR installations would be effectively independent of emission sources. CCS is installed directly onto power plants, etc. but that's classed differently.

Response:

Yes. The CDR installations that can operate effectively independent of emission sources are often referred to as direct air capture (see our recent work: Hanna et al., 2020).
CCS is the more "traditional" tech that would operate mostly in a high concentration environment in the power plant (e.g. to be combined with biofuel energy sources, BECCS).
No change is made to the text.

89 – worth comparing that to the "dry-gets-drier" pattern of global warming

Response:
Yes. We now stated the sulfur injection response is opposite to global warming.
"Based on the analysis from Simpson et al. (2019), the precipitation in tropical and extratropical regions shows a dry-get-wetter, wet-get-drier pattern due to the aerosol induced stratospheric heating, opposite to the well-known pattern of dry-gets-driver response due to global warming."

Sect. 2.1 This is a very short model description, and includes only one citation to one of the model versions used. Please elaborate.

Response:
This is now greatly expanded as suggested by Reviewer #1 as well, particularly to note the model difference and impact on the results.

Sect 2.2 Both experiments need to be described in more detail.

Response:
This is now expanded with an opening paragraph to state the objective of two experiments upfront, as well as acknowledging the "carbon capture" also includes some form of emission cut.

30 – "to stabilize TEMPERATURE at 2020 levels. . ." would be more accurate.

Response:
Yes. "climate"->" temperature".

35-41 – Is this based on another RCP, How large are the negative emissions, are there any differences in the land surface (more forest cover, etc.)? This description leaves out crucial details. Please outline them here even though they are explained in detail in the Sanderson paper.

Response:
It's a special scenario constructed by a group of us. We now added "(hereafter referred to simply as "Carbon Capture"), which is constructed to have a more aggressive decarbonization pathway than RCP2.6 (Sanderson et al., 2017)."
We also added in Section 2.2 that "Land cover and land use in these two pairs of simulations is consistently the same as in RCP8.5. "

Sect 2.3 – This is too little material for a sub-section, so I'd suggest cutting it or else expand it to discuss more of the analysis approaches employed, e.g. the time-periods covered and the procedure to normalize by global temperature change that is employed later.

Response:
The suggestion is taken. We now merge it into the opening paragraph of Section 2.2.

65-59 – This seems out of order, I'd suggest moving figures 1 and 2 out of the methods section and into the results section.

Response:
Although uncommon, we'd like to retain the first mentioning of these two figures in the method section, mainly because they are secondary and confirmatory and do not provide direct results.

Figure 1 – It would be much clearer to color one axis red and the other blue and use dashed versus normal to separate the experiments, then one could read the figure at a glance rather than having to get half-way through the description to know what is shown.

Response:
Revised as suggested.

Figure 1 – This shows ET and PET increasing in lock-step but Swann et al. 2016 (www.pnas.org/cgi/doi/10.1073/pnas.1604581113) showed all climate models they investigated diverging with PET rising rapidly and evaporation not rising at all. Swann et al. argued that this was due to the direct physiological effect of CO_2 suppressing transpiration. That would suggest that there is no direct physiological effect in your simulations, is that correct? If not, can you explain the difference between your results and those of Swann?

Response:
The physiological effect of CO_2 is included through the CESM and WACCM simulation but appears to be weaker than other models and the suppression effects are only strong over tropical forest regions.
See the response to general comments as well.

Figure 2 – "significant" – please elaborate here or in the methods.

Response:
Added. "where the differences are significant at 95% confidence level following the student's T-test"

L80-84 – These descriptions are incomplete. I cannot tell whether this is a reasonable approach as the terms are not defined.

Response:

We now remove the equation due to undefined terms. Instead, we further clarify in the text.

89-99 – This paragraph is hard to follow, I'd suggest revising it.

Response:
We now revised it thoroughly to bring clarity. It's now shortened to be "The same bias correction is applied to future simulations (CESM_RCP8.5 and CESM_CarbonCapture, in 2006 and afterward), which branched from the CESM_Historical in 2005. For WACCM_RCP8.5 simulations, we indirectly adjust them to agree with the corrected CESM_RCP8.5 in 2010-2019 ("present-day"). We could not adjust WACCM directly based on observations because the WACCM-RCP8.5 simulation only starts in 2010 and overlaps with the observation record (ending in 2019) for less than ten years. Similarly, WACCM_SulfurInjection was adjusted to match with WACCM_RCP8.5 in 2020 (when it branches from WACCM_RCP8.5). "

89-91 – Is the same bias correction applied to both models or are the separate historical bias corrections applied to the future in both models.

Response:
Bias correction is separately done to these two models because they apparently have a different historical bias (Fig 3 a and b).

93-94 – I don't believe this is the same "present-day" as in the GLENS experiments which I believe aim to keep conditions fixed at their 2010-2030 levels.

Response:
No. "present-day" is here defined to be earlier (prior to 2020) so that there is no difference between these simulations at present-day (Fig 2c).

Section 3.1 – The sub-section title seems to be at odds with the section title, perhaps change the section title?

Response:
You are right. We change the section title to be "Mitigation at the global scale".

Figure 3 – What is "major land"?

Response:
See the responses earlier and Table 2 caption.

Fgiure 3 – The change in axis range between b and d should be avoided. You should make clear visually or in the text that WACCM's precipitation over "major_land" is ~25% too high. This is huge!

Response:

Different models have different climatology. That is why the community showed anomalies for CMIP5, not raw output. We now make a note about the P bias in both models. "For example, comparing Fig 3b and d, WACCM appears to have a 25% positive bias for precipitation over land, and CESM has a smaller positive bias."

10-11 – What is WACCM's climate sensitivity? If it's not known then I'd reverse the order and highlight that it's higher than the already-high 4 C of CESM.

Response:
We added WACCM climate sensitivity now.

13-14 – This is the first time that this has been mentioned, I'd suggest highlighting this fact in the methods section.

Response:
Yes. We now mention the climate sensitivity of both models in the method section (the end of Section 2.1).

16-20 – The off-tropics injection is irrelevant as if the sulphates were less effective more would have been injected. Of course there's also the afore-mentioned high climate sensitivity, which should be mentioned.

Response:
Good point. Since the goal is to stabilize the warming in 2020, the cooling can be enhanced with more injection to achieve that. The paragraph is rewritten as :
"The two types of mitigation efforts, by design, would lead to a similar amount of temperature stabilization to a level close to the present-day. 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, some from locations off the tropics (Tilmes et al., 2017; and Kravitz et al., 2018)."

22 - "global major land" needs to be defined.

Response:
The definition of "major land" is the land regions over 60ºS to 60ºN. See Table 2 caption for rationale.

25-28 – Isn't it better to describe this as a fundamentally different experiment given their different aims? The carbon capture and GLENS experiments have different ends that end up producing a roughly similar temperature response, i.e. little change from 2020.

Response:
With respect, we disagree they are fundamentally different experiments because both are designed to produce a stabilized climate that's not too far from 2020, with different means though.

Please see the new opening paragraph of Section 2.2.

Table 2 – Seems unnecessary, suggest cutting. You've already shown this in figure 3, and reported many (perhaps too many) of these figures in the text.

Response:
We want to keep Table 2 in order to support the quantitative statements in the text. Without Table 2, it's hard for us and the readers to keep track of the numerical figures reported in the text.
Without Table 2, it would also be hard for reviewer #2 to back out the temperature changes between now and the end of the 21st century.

29-31 – These lines are unclear, larger than what? Also unclear what is being referred to. Is column3, mid-century P or is it all of the PET results? Do the variables count as a column?

Response:
Sorry for the confusion. There was no Column 3 in Table 2. We were mistakenly referring to an earlier version of Table 2 in which global results and major land results were presented separately. We have removed the statement in question.

35 – I'd avoid the term "mitigation" given its use to refer to emissions cuts in the climate literature. I'd suggest renaming this "mitigation potential" to something else.

Response:
I think climate engineering, in addition to emission cut, is also part of the mitigation/solution package.
We change "mitigation potential" to "potential mitigation benefits".

37 – explain that this is a reduction of a precipitation increase not a 100% reduction in precipitation.

Response:
Thanks, we corrected it.

38-63 – These paragraphs on the global-mean precipitation response ought to be revised, they are not well written. It is also strange to refer to an almost perfect reversal of RCP8.5 precipitation trends as being over-effective.

Response:
There might be some misunderstanding here. The over-drying is referring to other sulfur injection experiments previously published. For this model version, the over-drying is dampened.
The paragraphs are revised to strengthen the argument.

L63 – Could you state what "this precipitation-centered argument" is or else reframe this.

Response:
It's explained in the sentence immediately following: "precipitation alone does not reflect the full effects of the hydrological cycle on the terrestrial ecosystem.". We also change "this precipitation-centered argument" to be "this precipitation-centered perspective in assessing terrestrial aridity".

L67 – I'd avoid referring to sulfate injection as a mitigation scheme.

Response:
We change the "mitigation scheme" to "climate engineering schemes".

L67 – why aspire to do it and not just get on with it?

Response:
We change "we aspire to include " to " we also included".

70 – does it? I see a drying in all regions in P/PET and a wettening trend in most places except the amazon and central America.

Response:
We change " a close agreement" to " a broad agreement".

73-75 – this sentence is mangled.

Response:
Thanks for catching that. Rewritten as "Despite an increase in P, the projected PET growth that approximately scaled with T increase will exacerbate future land aridity".

75-77 – is this in the RCP8.5 experiment?

Response:
Yes. Added.

77 – "bring forth the benefit of curbing aridity worsening"

Response:
Revised to be "reduce the tendency of worsening aridity".

89 – I'd leave this type of commentary until the discussion as its tangential to what is being described.

Response:
Commentary language "(presumably shifting the climatic zone to a category beneficial to the cropland)" is now removed.

95-96 – Better to describe this as changing the sign of the trend as reversing is ambiguous in this context, i.e. it could mean simply offsetting.

Response:
Great point. We now made it clear that the drying trend is flipped to be wetting, instead of being "reverse". It now reads:
"Sulfur Injection appears to have this additional "benefit" of flipping the sign of the drying trend as projected in the baseline scenarios"

98-99 – the shorter response time doesn't explain this. P/PET rises rapidly then stops rising, Why is this?

Response:
The entire "on the timing" paragraph is rewritten, with the Fig 4 replotted. The P/PET growth rate becomes smaller in the second half of the 21st century because the P decreases at a faster rate than PET decrease, which again is due to the short response time nature of Sulfur Injection.

Figure 4. Again, these experiments are so different in character it seems odd to compare their time evolution. They have a different temporal evolution of forcing (in ways not described or explored in this paper) so its unsurprising that they have different temporal evolutions of climate response.

Response:
Thanks for asking because this is a key figure with the differences bringing new insights. The different temporal evolution of forcing in (f) is the main point we want to address because the carbon capture, despite having an earlier adoption in the effort (e), lags behind the Sulfur Injection.

Section 3.4. I'd avoid the generic sub-section title "summary" and give something specific to this section.

Response:
We change the generic sub-section title to "3.4 Summary: P/PET as the aridity metric and the faster/stronger benefits due to Sulfur Injection". We will follow the editorial guideline if this turns out to be too specific.

10-14 – As I mentioned earlier, the results in figures 1 and 2 do not show the same pattern for the different variables so I don't think it fair to say P/PET stands in for all measures of aridity as this text implies.

Response:
It's never our intent to claim P/PET stands in for measures of aridity. We select it to avoid getting into the weeds of too many metrics. We have now made it clear that Fig 1 and 2 did reveal subtle discrepancies especially at a regional level, and thus more metrics should be

included if the goal is for a comprehensive drought condition assessment as opposed to the comparison of two high hypothetical scenarios.
We added "Note that we select P/PET among other drought indicators to avoid the complexity of the analysis because the main goal of our analysis is to compare highly extreme mitigation scenarios. Fig 2 revealed regional discrepancies that suggest for a more comprehensive drought assessment, more metrics should be included in addition to P/PET."

18 – where is the CO2 captured results?

Response:
The mass captured/cut is shown in Figure 4e, a key piece of information.

21 – This type of discussion should be saved for the discussion section.

Response:
Since we do not have a dedicated Discussion section, and the discussion in the Conclusion section is mostly related to the limitation of the current study and implication for future studies. We still keep this quick comment here.

Section 4, how about "hydrological change in the Americas"?

Response:
Point taken. We change it to "4 Regional responses in the Americas"

Figure 5. This figure has too much going on and should be split or else the mid-century results cut.

Response:
We now split Fig 5 into Fig 5 and 6, one for CESM and one for WACCM. We do not want to cut mid-century results because the contrast between the near-term and long-term is a key point of ours.

The WACCM RCP8.5 results are missing which are crucial for interpreting the WACCM GLENS – RCP8.5 anomalies plotted.

Response:
The split enables us to add WACCM_RCP8.5 results to the first column of the new Fig 6.

I'd suggest replacing the bias-corrected model results with the original model results. It is unsurprising that the bias-corrected results look very similar to the observations. The models simulated present day has a direct bearing on its projections for the future whereas the bias-corrected present day doesn't.

Response:

We had already shown the original model results in Fig 2, so we did not make the changes. We had also discussed the different biases of the two models and implications on projection.

I'd also suggest making the CDR and GLENS anomalies relative to present-day rather than RCP8.5.

Response:
The anomaly in relative terms (%) is indeed calculated with the present-day as the reference. i.e, (CDR - RCP8.5) / present_day(2010-2019).

Finally, The caption format shouldn't include paragraph breaks.

Response:
Of course. Just suppose to make a reviewer's life easier since it's a busy figure. Will correct once sent out for the press. Thanks.

43-45 – which sees this to a lesser extent? Is the global value "major land"?

Response:
Clarified as although to a less extent "for the latter".
Change "global value to " Responses over global major land"

46-48 – this seems redundant

Response:
We shortened and merged it with the previous paragraph.

Table 3. The formatting / caption could be clearer here. Is CESM historical the same as historical after bias-correction, are they similar?

Response:
Yes. CESM_historial is after bias correction. Dryland area is based on absolute definite, i.e., AI < 0.65. That means we have to correct the bias. otherwise, we cannot obtain the dryland area.

51 – make clear that you are referring to the global arid area.

Response:
Global added.

4.2. This whole South America section is poorly written, I'd suggest revisiting it.

Response:
Suggestion well taken. Many revisions are included.

61-62 – revise this sentence.

Response:
Revised to be " While the changes over North America are largely consistent with the global land change, South America's responses are more complicated."

65 – I'd suggest using 2 significant figures here. What does the 15-30% refer to?

Response:
We did not go further to two significant figures because of the inherent uncertainty. 15-30% refers to relative change as shown in Figure 5.

66-68 – Is the amazon really the only region in the Americas to see a decline in precipitation in RCP8.5?

Response:
Yes. Original Figure 6b shows that North Brazil has a significant precipitation decline under RCP8.5. The decare over Easter Us appears to be less significant.

Figure 6 – Given the differences in the climates of the regions under investigation would it not be better to report results in percentage change terms? There's a similar ~50 mm/day increase in PET in West US and Brazil but presumably they have very different absolute values.

Response:
Good suggestion. But the percentage change terms are already reported in Table 2 and Figure 5. Also, we do need to present the absolute value changes of P and PET to understand the P/PET change. Using relative changes of P and PET to derive P/PET relative change would be misleading.

73-74 – this isn't a good description of what is in the figures, P is up in all regions in the CDR experiment which is just different from what is going on under RCP8.5 and around 80- 90% of the PET response is offset.

Response:
We are talking about Northern Brazil here which has quite different P responses compared to other regions examined.
The revised sentence is "This is achieved by both offsetting the precipitation decrease over Northern Brazil, but more importantly, by offsetting the projected PET increase by 80 to 90 % (Figure 6 b,c).".

79 – why not open with the big picture for GLENS then address this exceptional response?

Response:

Unclear to us what the reviewer meant by "big picture". We are well into the section of South America which we mainly focused on North Brazil (Amazon).

83-85 – it's not consistent with Simpson et al if it's the same data, it's just the same thing.

Response:
Wording changed to be "This additional suppression of tropical rainfall due to Sulfur Injection is also previously reported in Simpson et al., (2019),"

94 – continues? When did it start?

Response:
Revised to be "Unlike Sulfate Injection, Carbon Capture is consistently capable of mitigating the projected drying trend almost entirely over South America (similar to the global and eastern US cases)."

96 – There's a larger reduction in PET in GLENS, so the differences should be due to the differences in the precipitation response. Should probably note that the precipitation reduction in CESM-RCP8.5 is around twice as large as in WACCM-RCP8.5.

Response:
I think the reviewer misread the sentences as Amazon ("the precipitation reduction in CESM-RCP8.5 is around twice as large as in WACCM-RCP8.5" as shown in Figure 6b), while it's referring to sub-Amazon.
No changes are made except minor wording clarifications.

00-01 – Which regional trend is being referred to here?

Response:
"over the sub-Amazon" added and the paragraph merged with the previous one on carbon capture.

11-13 – Again, what is being referred to here?

Response:
It's about "sub-Amazon" over South Brazil.

Section 4.3. Again, I'd avoid using "summary" for a sub-section title.

Response:
We change it to "4.3 Summary: the unique responses over Amazon (North Brazil)", pending editorial approval.

37 – it's not a further decrease in precipitation if the decrease in precipitation is less than in RCP8.5.

Response:
It is indeed a further decrease in precipitation because the decrease in precipitation is more than in RCP8.5, as shown in the original Figure 6b.

50 – again, I'd suggest avoiding using mitigation in this way: "mitigation capacity"

Response:
We change it to "mitigation benefits". We do not see a problem calling Sulfur Injection a mitigation approach.

52 – Presumably WACCM-CDR (if you had run it) would be about 20% less effective than CESM-CDR by this measure as WACCM has a higher climate sensitivity and so the response would be divided by ~6C rather than ~5C.

Response:
We do not entirely agree. If WACCM_CDR were used, the <u>actual response</u> will also be different than what's in CESM-CDR, so it's not straightforward to claim that the normalized changes would be 20% less effective.

Figure 7. I don't see any value in including this figure. How is the statistical test applied here?

Response:
Figure 7 (now Figure 8) of the normalized change is a key figure to bring up the physical insights and we had retained it. The scaled change of P/PET induced by carbon capture and sulfur injection can tell the efficiency of different forcing.
The statistical test applied is the same as in Figures 2 and 5 so we did not repeat the details here.

56-59 – This is a result of not changing the scales of the plot to reflect the fact that you've divided it by ~5.

Response:
Correct.

62 – Higher climate sensitivity is the obvious driver of this but isn't mentioned.

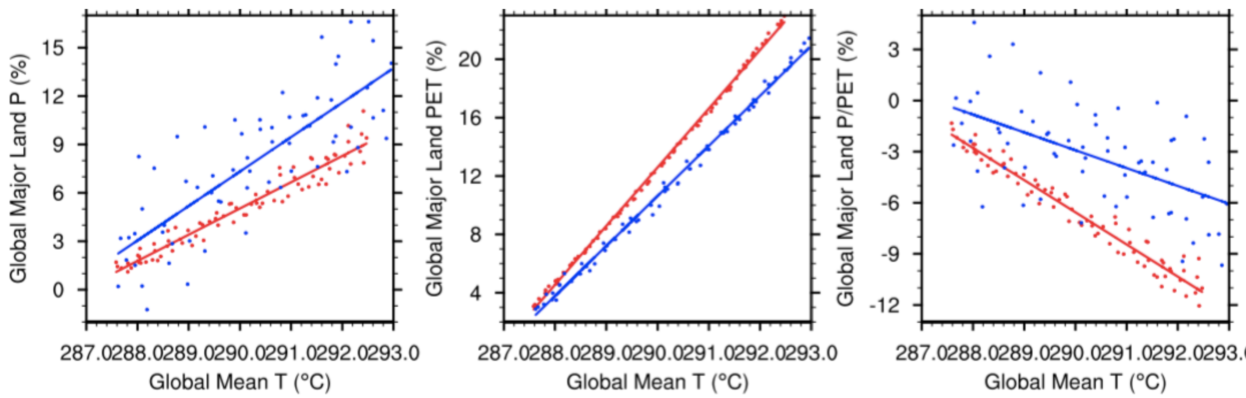
Response:
With respect, we do not think higher climate sensitivity is in play here because we have normalized the change with respect without global temperature change. This seems to be related to the confusion as in the general comment. In other words, the climate sensitivity will affect the spread of the points in Figure 8 (blue vs. red), but not much the slope of lines.

Figure 8. It is not clear what is going on in this plot. Which model is plotted? Are these the bias-corrected results or not? This plot needs to labeled more clearly.

Response:
As labeled and explained in the caption, Red for carbon capture (using CESM) and blue for sulfur injection (using WACCM), both bias-corrected.

I'd suggest producing a similar plot for the RCP8.5 results of the 2 model versions as I suspect that they could look quite different and this might be driving some of the difference attributed to the forcing differences here.

Response:



We plot the RCP8.5 results as suggested. There are some differences between two models, so we agree with the reviewer that model difference is partially contaminating the normalized results. We have fully acknowledged this caveat in Section 5 and also in the conclusion section and called for future research to systemperically address this issue (hopefully using multiple models).

85-87 – missing citation to earlier work.

Response:
Thanks, we added Lin et al., (2016a).

90-91 – missing again.

Response:
Thanks, we added Lin et al., (2016a).

95-97 – it could also be that tropospheric aerosols are concentrated over the land.

Response:
This is a great point. We now added, "...the stratospheric aerosols can be less effective in changing surface energy balance than tropospheric aerosols, which can induce warm cloud changes and are more concentrated over the land."

06-07 – rather than reversing, I'd recommend "changing the sign of" or some other construction that's less ambiguous.

Response:
We change "reversing" to "flipping". We also checked other places for consistency. This is a great suggestion.

Table 5 – The formatting is not great, I don't think the X / Y format is the best choice.

Response:
We struggle to find a better alternative without making it too complex.
Also, the format was used in Table 3 as well.

The simulated PET value should also be reported as well as a total or mismatch column.

Response:
The total change in PET in (%/°C) is shown in figure 8 and quotes extensively in the text already.

18-22 – This paragraph should be revisited.

Response:
It's rewritten as:
To summarize, the weaker sensitivity of Sulfur Injection globally, and the weaker response in absolute values over the Amazon region as shown in Section 4, are the two major counterarguments against the effectiveness of Sulfur Injection. The comparison with our earlier studies on tropospheric SO2 and volcanic eruption using different model configurations (the 20th century or Last Millennium; CESM with 1º or 2º resolution) demonstrate qualitative robustness of the results and the common physical mechanisms.

26 – this is not a counter-argument, P/PET is a common measure of aridity P is not.

Response:
We remove the "counteragrument" statement.

29-33 – The experiments are very different, this paragraph describes them as if they had the same goals.

Response:
As in response to general comments, we do think they are comparable because they share the same goal of stabilization climate at a predetermined level.

35 – I'd suggest "less effective at offsetting the amazon drying"?

Response:
Great. We change reverting to offsetting.

40-47 – This difference in climate sensitivity also undermines the normalization procedure in section 5.

Response:
We do not dispute that the difference in climate sensitivity is still perplexing the normalized results, but the normalization procedure is specifically conducted to mitigate the caveat of model difference.