Response to reviewer #2

Review:
**The submitted paper documents the experimental design for the CDRMIP
suite of experiments, designed to explore model uncertainties in Earth
System response to climate engineering through potential anthropogenic
removal of carbon dioxide from the atmosphere. The MIP is well motivated,
and the introduction does a good job of framing why such a MIP would be
useful.
The paper should certainly be published, and I look forward to seeing the
results of the MIP. I have some minor comments only, which I attach for the
authors' consideration.**

Minor Comments:

**1. The details of the experimental design need clarifying in places. For
example, a number of the experiments require 'constant forcing' for non-
CO2 agents, but the authors do not explicitly state how to implement this.
Should aerosol concentrations be held constant, or should emissions be
held constant?**

*Sorry for leaving out these details. We have added a paragraph to Section 4 to
clarify what we mean by constant forcing. This paragraph reads,*

"In some of the experiments described below we ask that non-CO_2 forcing (e.g.,
land use change, radiative forcing from other greenhouse gases, etc.) be held
constant, e.g. at that of a specific year, so that only changes in other forcing, like
CO_2 emissions, drive the main model response. For some forcing, e.g. aerosol
emissions, this may mean that monthly changes in forcing are repeated
throughout the rest of the simulation as if it was always one particular year.
However, we recognize that models apply forcing in different ways and leave it
to individual modelling groups to determine the best way hold forcing constant.
We request that the methodology for holding forcing constant be documented
for each model."

**2. There is almost no consideration of internal climate variability,
recommended ensemble size, and what role that might have. How many
ensemble members are required for each of the experiments to assess the
desired signal? If it is only 1, can the authors demonstrate that a single
simulation can produce a sufficiently significant result to differentiate the
structural differences between different models in the presence of climate
noise?**

*We do recommend that groups conduct 3 ensemble members (Section 3.3) to deal
with variability. However, for CDR-MIP, interannual variability is likely to be a*

larger issue than internal model variability. Pervious studies such as Hewitt et al., (2016) *that looked at this issue with a focus on the carbon cycle, which is especially relevant for CDR-MIP, found that when comparing simulations of CMIP5 scenarios for land-carbon fluxes, the model spread was so big that it was the primary source of uncertainty. While for ocean carbon uptake, the variance attributed to differences between representative concentration pathway scenarios exceeded the variance attributed to differences between climate models. In most models "internal variability" (assuming this means "sensitivity to perturbed initial conditions") was fairly small – especially on decadal scales. Interannual variability of carbon fluxes was high, but tended to even out on >5 year timescales. Based on this knowledge, we recommend that modelling groups perform at least three ensemble members to reduce this uncertainty related to variability, but leave it up to each group to determine how much of an issue this is and whether it requires more or fewer runs. Thus, section 3.3 states that,* "We encourage participants whose models have internal variability to conduct multiple realizations, i.e. ensembles, for all experiments. While these are highly desirable, they are neither mandatory, nor a prerequisite for participation in CDR-MIP. Therefore, the number of ensemble members is at the discretion of each modeling group. However, we strongly encourage groups to submit at least three ensemble members if possible."

3. It isn't clear how a proposed experiment esm-ssp534-over differs from the existing C4MIP ssp534-over-bgc. Could the authors make this more clear?

The reviewer is likely referring to the statement in section 4.2 where we stated that,

"We also highly recommend that groups conduct the ScenarioMIP *ssp534-over* and *ssp534-over-ext* and C4MIP *ssp534-over-bgc* and *ssp534-over-bgcExt* simulations as these runs will be invaluable for qualitative comparisons."

We agree that the relationship between these simulations was not clear from this isolated statement. We have deleted this statement to avoid repetition (as recommended by reviewer #1) and now highlight the relationship between these simulations in Section 3, where more detail is provided. Here we state that:

"The C4MIP experiment *ssp534-over-bgc* is a concentration driven "overshoot" scenario simulation that is run in a partially coupled mode. The control run required for analyses of this simulation is a fully coupled CO_2 concentration driven simulation of this scenario, *ssp534-over,* from the Scenario Model Intercomparison Project (ScenarioMIP). The CDR-MIP experiment, *C2_overshoot*, which is a fully coupled CO_2 emission driven version of this scenario, will provide additional information that can be used to extend the analyses to better understand climate-carbon cycle feedbacks."

We also have similar statements in Section 3.2 that read,

"We also highly recommend that groups run these additional C4MIP and

ScenarioMIP simulations:

- The ScenarioMIP *ssp534-over* and *ssp534-over-ext* simulations, which prescribe the atmospheric CO_2 concentration to follow an emission overshoot pathway that is followed by aggressive mitigation to reduce emissions to zero by about 2070, with substantial negative global emissions thereafter. These results can be qualitatively compared to CDR-MIP experiment *C2_overshoot,* which is the same scenario, but driven by CO_2 emissions.

- The C4MIP *ssp534-over-bgc* and *ssp534-over-bgcExt* simulations, which are biogeochemically-coupled versions of the *ssp534-over* and *ssp534-over-ext* simulations, i.e., only the carbon cycle components (land and ocean) see the prescribed increase in the atmospheric CO_2 concentration; the model's radiation scheme sees a fixed preindustrial CO_2 concentration. These results can be qualitatively compared to CDR-MIP experiment *C2_overshoot,* which is a fully coupled version of this scenario."

4. Could the authors expand on what processes would result in yr2010co2 differing from esm-hist-yr2010co2-control, given that if compatible emissions are correctly diagnosed, they should be identical? The only case, to my mind, where this would not be true is if internally-generated climate noise was capable of changing the compatible emissions requirements. However, if this is the case, then the experimental design is insufficient - and an ensemble of yr2010co2 simulations would be required in order to assess the central estimate for compatible emissions.

In the test simulations that we have performed with both an ESM and EMIC it appears that climate-carbon cycle feedbacks, which become evident when atmospheric CO_2 is allowed to freely evolve, can result in the diagnosed CO_2 emissions forcing either slightly under- or overestimating the emissions needed to reach 389ppm. We agree that in such cases our original design was insufficient and have added text to clarify the necessary steps to achieve the correct atmospheric CO_2 concentration. This text reads,

" If there are significant differences, e.g., due to climate-carbon cycle feedbacks that become evident when atmospheric CO_2 is allowed to freely evolve, then they must be diagnosed and used to adjust the CO_2 emission forcing. In some cases it may be necessary to perform an ensemble of simulations to diagnose compatible emissions."

5. In esm-hist-yr2010*, what RCP/SSP should be used if 389ppm is not reached during the historical period?

For groups performing the CMIP6 historical simulation achieving 389ppm should not be a problem as this is part of the prescribed historical forcing. However, we agree that it could be an issue for those using a CMIP5 model configuration and forcing. We have therefore recommended that they use the RCP 8.5 simulation to

reach 389 ppm and the sentence now reads, " An existing run or setup from CMIP5 or CMIP6 may also be used to reach a CO_2 concentration of 389ppm, e.g., the RCP 8.5 CMIP5 simulation or the CMIP6 *historical* experiment."

Typos/presentational points:

Line 50: comma after climate

Corrected.

Line 118: Do any of the 2 degree scenarios (which have not already diverged from historical emissions) require no CO2 removal? I'm not aware of them. Could they be cited?

We are not aware of any of limited warming scenarios without CDR either and have changed the text accordingly. In our original statement we had been referring to scenarios that have already diverged from historical emissions, but now realize that it doesn't make sense to refer to them.

Line 126: suggest "are not yet a commercial product"

Change made.

Line 395: This paragraph seems to imply that a GCM can inform policy which differs only in terms of the relative sizes of positive and negative fluxes which make up a net anthropogenic flux. This seems to be true only for a subset of CDR approaches where there are long term consequences of removal for future fluxes (e.g. reforestation), but not really for direct air capture. Perhaps this could be clarified

We have clarified this statement to address the issue raised here. The sentence now reads, " This relates to the policy relevant question of whether in a regulatory framework, CO_2 removals from the atmosphere should be treated like emissions except for the opposite (negative) sign or if specific methods, which may or may not have long-term consequences (e.g., afforestation/reforestation vs. direct CO_2 air capture with geological carbon storage), should be treated differently."

Line 464: Suggest using a word other than "control" here, which is almost universally interpreted as a constant forcing simulations in other CMIP6 MIPs.

Done. "control" has been replaced with "simulation".

Line 971: Is esm-535-over-ext a typo?

Yes, this is a typo and has been corrected.