

Response to reviewers

Thank you to the reviewer for your detailed review of our manuscript. We have responded to each comment in full and outlined the changes we will make to the manuscript to address your comments in this document. Our responses are in black font in response to review comments in blue, and where we quote new text, this is in italic.

Reviewer comments

Overall, this is a very well written paper that clearly and accurately documents the experimental protocol for the third phase of the Reduced Complexity Model Intercomparison Project (RCMIP). This is valuable because RCMIP is an important contribution to the understanding and improvement of reduced complexity models, which themselves are a valuable tool for both theoretical climate science and applied climate policy analysis.

I encourage the journal to accept this paper after a number of minor revisions (see following list). The most substantive comment is a criticism regarding the use of SSP3-7.0 for the CH4 variant experiments because of its unusual nature in terms of non-GHG emissions. The only other comment that potentially rises beyond just wording changes is a request to include tropospheric ozone concentration (not just forcing) as an output. This has potential wider implications about including links between global average O3 concentration and carbon cycle behavior, which I think would be consistent with the drive of RCMIP3 to improve carbon cycle representation and understanding.

Thanks for your comments, we respond to each comment in turn here and in the revised manuscript.

Lines 1-29: The value of RCMs is a good starting point for this paper, and this is a pretty good description. However, I think that the authors have missed both a reason that RCMs are valuable, and a key application:

- As I state in Sarofim et al. 2021 (<https://pmc.ncbi.nlm.nih.gov/articles/PMC8311623/>), RCMs have at least three key advantages: “(1) computational efficiency, (2) adjustable model parameters and (3) lack of chaotic behaviour.” The drafted text addresses #1 and #2, but not #3. #3 is particularly important for understanding the response of the climate system to small perturbations (e.g., emissions pulses as in the SC-GHG – see next comment, or analyzing the benefit of a specific policy or regulation)
- RCMs are key for estimation of the Social Cost of Greenhouse Gases: I would suggest citing Rennert et al. 2022 (<https://www.nature.com/articles/s41586-022-05224-9>) and/or the EPA 2023 SC-GHG report or the NAS 2017 SC-GHG assessment.

We agree this is a valuable characteristic of RCMs that enables important uses, particularly in the policy sphere. The first paragraph in the introduction now includes the following text to acknowledge this:

Moreover, the absence of chaotic behaviour in these models enables the evaluation of climate impacts arising from comparatively small forcings that would otherwise be undistinguishable from interannual variability in ESMs. For instance, RCMs are a key tool in the calculation of the

Social Cost from additional emissions of greenhouse gases (GHGs, Rennert et al., 2022; EPA, 2023).

Line 51: I think this is the first time flat10MIP is mentioned. A 1 sentence definition would be helpful for those readers (such as myself!) who don't immediately recognize the term – I might suggest a parenthetical like (“a series of idealized experiments starting with constant 10 GtC/year CO2 emissions”)

We agree a clarification would be helpful. The text now reads:

The increasing focus of CMIP7 on emissions-driven simulations (Sanderson et al., 2024), including the new idealized experiments from flat10MIP (Sanderson et al., 2025) which feature a constant CO2 emissions rate of 10GtC/year, provides a direct bridge to the native way of operating RCMs.

Line 95: incomplete WMO reference

Thank you. Now corrected in the text.

Line 125: I just want to commend the RCMIP team: I think this experiment will be very helpful for disentangling feedbacks (I could commend the team for much more - as I said, this paper is very good - but this experiment had particular appeal to me)

Line 135: And again, I think this will be very important. Dietz et al. 2021 (<https://www.journals.uchicago.edu/doi/abs/10.1086/713977>) was a great paper showing that modern RCMs behaved very different from the climate modules embedded in the 2010 USG SC-GHG IAMs, but I think the lack of proper GCM/ESM pulse experiments for “truthing” is something that the community needs to address.

Thank you for your kind and encouraging words. We are very excited about analysing the data from these experiments once the submissions have come in from the modelling teams!

Line 140: I think this experiment is describing a process where:

- Each RCM back-calculates CO2 emissions from a 1 pct CO2 concentration driven experiment
- This back-calculated CO2 run is the set to zero at the point at which it reaches 750/1000/2000 PgC cumulative carbon.

I think that this can be written more clearly.

Thanks for pointing out this convoluted explanation. We have rephrased it to something that should make the nature of the experiment clearer to the reader:

Compatible carbon emissions with an esm-1pctCO2 experiment are back-calculated for each model, and subsequently set to zero after certain levels of cumulative emissions are reached (750, 1000 and 2000 PgC).

Line 165: If I've read it right, flat7.5-cdr and flat7.5-rev are very similar, with the difference being that cdr ramps from 7.5 to -7.5 over a hundred years, whereas rev zeroes emissions out for those hundred years. But the cumulative emissions from the two experiments should be identical everywhere except those 100 years. If this is true then:

- I would put these experiments adjacent to each other
- I would explicitly note this relationship, as the rationale for having 2 similar experiments is presumably to look at how the ramp versus immediate jump of emissions lead to differences.

Your interpretation of the experiments is correct. We agree that placing them adjacent to each other is a good idea. We have also updated Figure 2 to reflect the new order. The main reason for this new experiment is to allow a cleaner evaluation of the transient response to cumulative negative emissions with a period with zero emissions between the two regimes. We have now updated the text to make this more explicit. That part of the article now reads as follows:

– *esm-flat7.5-cdr, esm-flat10-cdr, esm-flat20-cdr*: Following an initial 100-year period, emissions are linearly reduced over a 100 year period to the negative of the original flat rate (e.g., from +10 GtC/yr to -10 GtC/yr in *esm-flat10-cdr*) to assess the system’s response to large-scale CDR. Then, after 300 years when cumulative emissions reach zero, emissions are abruptly set to zero.

– *esm-flat7.5-rev, esm-flat10-rev, esm-flat20-rev*: Following an initial 100-year period, emissions are held constant at zero for another 100-year period, then reversed to the negative of the original period for another 100 years, and finally abruptly set to 0 until the end of the experiment. This is a new addition with respect to the original flat10MIP protocol, similar to the *esm-flat*-cdr* experiments but with an abrupt emissions cessation rather than linear. The objective is to allow a cleaner assessment of the transient response to cumulative removals by introducing an interim period without external forcing between the positive and negative emissions regimes.

182: I don’t think that the source of historical concentrations has been described before this point. I would like a shorthand that could be used here and going forward like “Foster24” in the same way that “CMIP6” describes the alternate concentration source (you may need other references for the historical emissions source, volcanic/solar forcing source, etc...)

These are just the standard CMIP concentration-driven historical experiments. There is no single publication to reference for CMIP7 datasets, so nothing like “Foster24”. We listed all the relevant sources of data at the end of section 2 (just before the start of subsection 2.1). We would like to align experiment names as much as possible to standard names to minimise future confusion, so we have retained the “historical” and “historical-cmip6” names (the latter is just a necessity to distinguish the two experiments). We have, however, updated the referenced text to make it clearer that these are just the standard CMIP experiments:

historical and historical-cmip6: Standard concentration-driven historical simulations from CMIP. The former uses CMIP7-era forcing agents and runs from 1850 to 2023, while the latter uses CMIP6-era forcing agents and runs for the 1850–2014 period.

187: thank you for doing isolated aerosol, GHG, and CO2 experiments. I wish that CMIP could do more of this too.

Thanks for the appreciation!

218: from/form typo

Thank you. Now corrected in the text.

222-230: I am concerned by the reliance on SSP3-7.0 for these experiments. See Shiogama et al. 2023 (https://ideas.repec.org/a/nat/natcli/v13y2023i12d10.1038_s41558-023-01883-2.html): as an outlier scenario it will introduce questions about the applicability of any results. This is particularly important for methane, as NO_x, CO, and VOC emissions interact with the hydroxyl radical, which leads to changes in methane lifetime, so using SSP3-7.0 for low and high CH₄ experiments seems a bit self-defeating. I know that SSP3-7.0 has the advantage of having a pre-made lowNTCF companion scenario, but RCMIP could take SSP4-6.0 and make its own high and low NTCF versions of it for the purpose of this exercise.

Thank you for raising this point. These scenarios were indeed selected to align with the MethaneMIP experiments, with the aim to maximise comparability between RCM output and existing ESM data. As a result, RCMIP3 has generally avoided designing any ad-hoc experiments unless a clear gap in coverage exists. While we agree that the outlier nature of SSP3-7.0 may raise some applicability questions, we believe this will be partly mitigated by the inclusion of complementary MethaneMIP experiments in our protocol (esm-allGHG-scen7-H-CH₄L, esm-allGHG-scen7-L-CH₄H, and methanemip-TM(+BC)-allGHG which branches off SSP2-4.5). Accordingly, the scenarios have been retained in their original form.

Line 227: typo: lowNTFC should be lowNTCF

Thank you. Now corrected in the text.

Line 230: SSP3-7.0lowNTCF is not “an otherwise low-forcing background”- it is only low in non-GHG emissions.

Thank you for pointing out this ambiguity/mistake. We have updated it now to be more precise:

This is designed to isolate the specific impact of a high methane pathway against an otherwise low non-GHG forcing background.

Line 238: I think that this is the first time that the CMIP7 Scen7-H/L scenarios are mentioned: it would be worth a quick description here.

We agree a quick description may be useful for readers. We added a brief summary of the scenarios in section 2.2.3, as we deemed it a more appropriate place to add that description. The text added was:

The scenarios encompass a wide range of plausible futures, from very low emissions (-VL) compatible with the highest possible ambition on climate change, to high emissions (-H) arising from a rollback of current mitigation policies. There is also a “medium” scenario (-M) similar to current policies, as well as variations of the previous scenarios with differing assumptions around policy strengthening and negative emissions (-HL,-ML,-VL, and -LN).

Table 2: As previously mentioned, I think a short modifier to “historical” (e.g., Foster24) would be useful in this table. Also, are the methanemip-TM-allGHG experiments all based on SSP2-4.5? If so, that would be useful to include in the description.

Thanks for your suggestion. As previously discussed, we preferred to keep the experiment names as aligned with standard experiment names as possible to minimise confusion concerning experimental design.

We agree that the fact that those two methane experiments are based on SSP2-4.5 would be useful to include in the table, so now the description reads:

methanemip-TM-allGHG -> Methane technological mitigation scenario based on SSP2-4.5 (all GHGs)

methanemip-TM+BC-allGHG -> Methane technological mitigation + behavioural change based on SSP2-4.5 (all GHGs)

Line 312: I think it should be ECS and TCRE, not “ECS, TCRE”

Thank you. Now corrected in the text.

Line 324: typo: “above” not “aboved”

Thank you. Now corrected in the text.

Line 352: there is an “air” where there should be an “ocean” in the 2nd variable

Thank you. Now corrected in the text.

Table 4 & Table 5 titles: should Table 4 be, “Atmospheric Composition: Major GHGs”, and Table 5 be, “Atmospheric Composition: Other GHGs”?

Thank you for the suggestion. We agree it is a better description of the table contents, so the titles have now been updated with the suggested names.

Table 5: I would like to put in a request for tropospheric ozone concentrations (not forcing) in the atmospheric composition readout. Not all (possibly, even, very few) RCMs will have it, and I recognize it is a very heterogenous substance so there are issues presenting a global average value, and yet, it is so important for ecosystems and carbon uptake that I think it would be a key improvement to include it. The GPP impacts of ozone could be on the order of 2-12% (see Unger et al. 2020). The inclusion of ozone/carbon cycle impacts will also matter a lot for looking at changes in methane emissions (because methane is such an important global ozone precursor) (see Collins et al. 2010). And it can even give a little insight into human health impacts.

Collins, W. J., S. Sitch, and O. Boucher. "How vegetation impacts affect climate metrics for ozone precursors." *Journal of Geophysical Research: Atmospheres* 115.D23 (2010).

Unger, Nadine, et al. "Mitigation of ozone damage to the world's land ecosystems by source sector." *Nature Climate Change* 10.2 (2020): 134-137.

Thank you for your suggestion. We agree that this addition may be of relevance for multiple uses, so we have included it in the protocol (and communicated it to modelling teams in the submission period). Mirroring the stratosphere/troposphere split in the original protocol for ozone ERF, we have now added two Tier 3 variables to allow the reporting of stratospheric and tropospheric ozone concentrations. This is also reflected in Table 5.

Table 6: Typo, N20 should be N2O (zero v. oh)

Thank you. Now corrected in the text.

Table 15: Four more occurrences of the N20 vs. N2O typo (it is a common error that has always been a pet peeve of mine, along with CO2 v. CO₂ – though I didn't see the latter one in the paper)

Thank you. Now corrected in the text. We also checked for any more instances of this mistake (and CO₂), but could not find any more.

Other changes

- The author list and acknowledgment sections have been updated to better reflect the contributions to the manuscript.
- The description of the hist-aer experiment has been expanded to provide additional guidance to modellers on how the experiment should be run. The description now reads as follows:
 - hist-aer, hist-GHG, and hist-CO₂: Single-forcing historical simulations driven by emissions of aerosols-only, concentrations of all GHGs and concentrations of CO₂-only, respectively. These are crucial for attribution studies. *To align with the original experiment design and better isolate aerosol impacts, the hist-aer experiment should be run without interactive ozone and methane chemistry modules, as some reactive gas species (e.g., CO and VOC) interact with these atmospheric gases.*
- The part of the manuscript that said input for the latest ScenarioMIP experiments may not be ready by the publication time of the protocol has been deleted, as these became available over the manuscript review period. This input has now been shared with modelling teams as part of the protocol.
- Over the manuscript review period, we made a minor update to the carbon cycle targets we include in our protocol to calibrate and benchmark RCMs. While the original protocol version used the land and ocean carbon sinks from the latest Global Carbon Budget (Friedlingstein et al., 2025b), the latest protocol iteration employs an updated version of those same figures following a new methodology (Friedlingstein et al., 2025a). This new methodology will be employed in future Global Carbon Budget publications going forward, as well as the upcoming IPCC AR7. Consequently, we deemed it necessary to update our protocol accordingly to align with these projects and maximise its impact. The description of the global carbon budget components in our target section (as well as Table 3 and Figure 4) has been updated to reflect these changes:
 - *Global Carbon Budget components: For models able to partition carbon into ocean and land reservoirs, RCM carbon fluxes should be consistent with the central estimates and uncertainties of the land and ocean carbon fluxes we provide. These fluxes are reported by a revised version of the 2024 Global Carbon Budget following an updated methodology (Friedlingstein et al., 2025a). This updated methodology will underpin future annual updates of the Global Carbon Budget and the upcoming AR7 assessment, and has therefore been included in this intercomparison to maximise its relevance in future assessments. RCMs will be evaluated against the 2014–2023 assessed budgets for the net ocean and land carbon sinks from this revised methodology.*
- There was a small, but important, mistake in Table 3 regarding the baseline period for the aerosol ERF target. The original manuscript stated this baseline to be the 1850-1900 period, when it should have said 1750. This has now been fixed. This was only a mistake

with the protocol paper, with the underlying data including this benchmark and published alongside the manuscript remains the same.

- An additional Tier 2 variable covering global heterotrophic respiration has been added. Our intention was to include this in the original protocol version, but its addition was overlooked.
- We have updated the reference list to include published versions of pre-prints where available.