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

Review of "Reduced Complexity Model Intercomparison Project Phase 3: Experimental protocol for coordinated constraining and evaluation of Reduced-Complexity Models"

This manuscript describes the effort to coordinate the RCMIP phase 3 project, including description of forcing, tuning, and standard model outputs. The work is very important for benchmarking simple climate models and will coordinate many simulations that will be considered for writing the next IPCC report. It is thus clearly an important contribution, it is suitable for publication, and this journal is appropriate for such an effort.

My comments are minor, the authors have achieved their goals and the paper is well written, clear, and concise. I recommend publishing the article after giving the authors a chance to make (optional) minor revisions.

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

Minor comments:

1. One major aspect of this work that could be potentially controversial is the proposal for a standardized approach to tuning RCMs. The authors use language that suggests the standardized approach is optional, which suggests that it is not expected to be a universally accepted proposal. Could this standardization have some negative impacts, for example because they may decrease some aspects of model diversity? Are these benchmarks provided with uncertainty? A bit more discussion on the pros/cons to adopting the standardized approach could provide context and strengthen the argument for why modeling groups should adopt this procedure.

   Thank you for your questions. Our main concern regarding the provided benchmarks for model tuning was the time availability of modellers to constrain their models (RCMIP is generally an unfunded activity). Hence, why, even if highly-desirable, the calibration was left as optional. This is briefly discussed in the text.

   We do not anticipate opposition to the idea of a common calibration approach based on observations. Generally, the need for a common calibration approach to reduce model differences arising from different tuning processes is understood by modellers, and all selected variables are broadly high-level historical quantities that RCMs should be able to reproduce anyway to some extent. Furthermore, a similar exercise was conducted in the AR6 to align with WG1 benchmarks and hand over a calibrated set of models to WG3 for scenario categorisation. We do provide timeseries with uncertainty ranges, so modelling teams are free to decide how and to what degree they include these data in their calibration stage.

We did receive some comments regarding the suitability of the data employed for these six climate variables, so we have made a small change for the carbon fluxes. More on this can be found in the "Other changes" section below.

2. Somewhere in the introduction it might be nice to add one paragraph with a brief background on RCMs (e.g., some physics discussion). There are some nice review studies already out there, and that is not the purpose of this document. But a little more background information may help ease the burden for readers who have less background on the topic.

We agreed that a brief paragraph with some background on the nature of RCMs would be useful for readers. We have added this text at the beginning of the introduction:

Reduced-complexity Models (RCMs), also known as Simple Climate Models (SCMs), are a highly parametrised variety of climate models designed to simulate Earth system dynamics in a computationally efficient manner. To achieve this, RCMs operate at low spatial and temporal resolutions, typically using global-mean annual averages, and represent many processes through parametrisation rather than explicit simulation. For instance, their temperature components commonly consist of zero-dimensional energy balance models that translate the imbalance between incoming and outgoing radiation into changes in global mean temperature. The nature and diversity of RCMs are discussed in greater detail by Romero-Prieto et al. (2026).

3. It would be informative to include comments on computational cost to run various tiers of the experiment set (probably a rough range given varying complexity) and of storage requirements for the tiers of outputs.
Thanks for the suggestion. While we agree such numbers would be interesting, assessing computational cost is a challenging task to perform. Particularly so when one considers the wide variety in model complexity and requirements, setups and number of ensemble members we expect in model submissions. Hence, we have preferred to not add such an estimation. Storage requirements are easier to estimate, however, based on current submissions. Accordingly, we have added the following line in section 4.2:
We estimate that a complete experimental submission consisting of several hundred ensemble members for a medium-complexity RCM would require several dozen gigabytes of storage space.

4. L70: You could comment on some statistics of participation in this effort, how many individuals responded, what is the representation of different groups, etc.

Thank you for the suggestion. We agreed some more detail about participation would be useful. We have added the following sentence to include that detail:

This allowed the co-design of the protocol with the wider RCM modelling community, incorporating input from representatives of multiple RCMs (ACC2, CICERO-SCM, FaIR, MAGICC, and OSCAR)

5. L87: This is the only place the reader is referred to as "you", maybe rewrite.

Thank you. Now corrected in the text to:
If running the full 1000 years is prohibitive, a higher number of simulations for a shorter time period is preferable.

6. L319: Can this pyrcmip repo be tagged to a doi/citation?

   Repo has now been archived to zenodo and a citation included in the text instead of the URL.

7. Table 8 & Table 10: If there are standard units it would be useful to include them here.

Thank you for the suggestion. Standard units are indeed defined in our protocol and are available in both the protocol spreadsheet hosted on the Zenodo archive and the supplementary material. We chose not to include them directly in Tables 8 and 10 because, for consistency, units would then need to be added across all tables, which we felt would unnecessarily increase their complexity and reduce readability for more casual readers of the protocol. Users who require this level of detail, either modelling groups preparing submissions or researchers intending to use the datasets, would in any case be expected to consult and download the full protocol documentation, where the experimental setup and variable definitions are described comprehensively. A reference to units is made in the first paragraph of section 4 and we also updated the “Code and data availability” to make an explicit reference to units:

The latest versions of the protocol – including requested experiments and variables with associated units – as well as the input dataset and the code to generate that dataset can be found at https://gitlab.com/rcmip/rcmip-phase-3.

Minor/grammatical

- The paper is mostly very well written, but there can be a tendency for some sentences to run-on a bit. I suggest a little editing to try to avoid this tendency (e.g., L22-29 is two consecutive (long) single sentence paragraphs that could be rewritten for clarity; L32, L52 also have long, multi-comma sentences that could be rewritten for clarity).

We agree some parts of the manuscript would benefit from streamlining the presentation of ideas. We have done some light editing to do so, without changing the content presented in the first version of the manuscript. For instance, the text now reads as follows for the two sections of the text mentioned in the reviewer comment:

Editing of L22-29:

RCMs have a long history of important scientific applications. Examples include assessing climate outcomes from Integrated Assessment Model (IAM)-derived emissions scenarios (Nicholls et al., 2022), direct coupling with IAMs (Baumstark et al., 2021; Stehfest et al., 2014; Hartin et al., 2021), providing non-CO2 adjustments to the remaining carbon budget (Lamboll et al., 2023), producing rapid answers in response to climate policy announcements (Bertram et al., 2020), and assessing uncertainties induced by volcanic forcing on climate projections (Chim et al., 2025), to name a few. However, the trustworthiness of RCMs relies on their simulation skill. They need to be able to adequately emulate the behaviour of more complex climate models (Smith et al., 2024), to appropriately reproduce large-scale indicators of observed climate (Forster et al., 2025), and produce plausible projections of past and future climate change (Smith et al., 2021; Verkerk et al., 2025).

Editing of L32:

RCMIP Phase 1 (Nicholls et al., 2020) provided a standardized protocol for generating the global-mean temperature projections for the IPCC's Sixth Assessment Report (AR6), demonstrating that the ensemble of RCMs was fit-for-purpose in reproducing historical warming and emulating the behaviour of the more complex models from CMIP5. This first protocol requested a single best-estimate ensemble member from each model provided, with the stipulation that this member should have an equilibrium climate sensitivity of 3°C.

Editing of *L52:*

RCMIP3 will leverage these synergies to provide rapid, probabilistic projections for CMIP7's "Assessment Fast Track" (AFT) experiments and to robustly assess key climate indicators like the Transient Climate Response to cumulative CO2 Emissions (TCRE) and the Zero Emissions Commitment (ZEC). These indicators are central to the latest assessments of the state of the climate system (Forster et al., 2021).

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 ad 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 CarbonBudget 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.