EGUSPHERE-2025-3604
The TIPMIP Earth system model experiment protocol: phase 1

Responses to Reviewer Comments

Dear Editor and Reviewers

We are grateful to both reviewers for their constructive comments and suggestions for our paper. These have greatly improved the manuscript. We have endeavoured to address each comment in this ***reviewer response*** document. Below, we reproduce the reviewer comments (**in black**), then add our response to each comment (**in red**). Where a block of text has been added to the manuscript, we highlight this in this document *(in red italics*).

Best wishes,

Prof Colin Jones

Reviewer 1

Thank you for inviting me to review the paper: "The TIPMIP Earth system model experiment protocol: phase 1" by Jones et al.

1. With climate science under intense scrutiny, papers like these are extremely important. Such a manuscript demonstrates that the research community is open and transparent.
   The role of description or methodological papers is just that, to set out proposed frameworks for simulations. And in that sense, the manuscript does well and fulfils the intended purpose. However, to provide some context and reasoning, it might be worthwhile to add a couple of sentences (at the beginning and in the discussion) explaining why there is such a rapidly growing interest in "overshoot" forcings. I would argue that there are four main reasons for modelling such scenarios.

- Society is likely to find many aspects of substantial climate change difficult to manage, leading to an eventual strong push to lower global temperatures.
- Society has set thresholds for global warming that it aims to stabilise at, notably around 1.5 or 2.0 degrees above pre-industrial levels. Since we are close to these benchmarks, achieving them will most likely involve a period of "overshoot".
- Studying the cooling period really helps understand the amount of negative emissions needed, raising challenges for others to question their feasibility (e.g., levels of geoengineering).
- If climate change triggers regional tipping points, then the second less-discussed issue is that of hysteresis. By definition, this might become an issue with "overshoot" scenarios, and so things are not reset for falling temperatures. The TIPMIP projections

described might be useful to drive offline impact models believed to potentially exhibit hysteresis.

Thanks for the suggestion. In the 2nd paragraph of the introduction, we have added the following text:

There is increasing interest in global warming overshoot scenarios (i.e. temporary exceedance of a global warming target that is returned to at some later date; Schleussner et al., 2024, Reisinger et al., 2025) and in the long-term response of the Earth system to net zero CO2 emissions at temperatures warmer than today (Allen et al., 2022, Frankhauser et al., 2022). Key outstanding questions include, the likelihood, timing, and magnitude of overshoot, the consequences arising from such an overshoot, the feasibility of cooling the planet back from a peak warming to a (stabilized) target level, the risk of hysteresis once a stabilized climate is achieved, and the long-term response of the Earth system to net-zero emissions at temperatures warmer than today.

Our protocol lends itself to addressing a number of these questions. Areas of particular interest include (i) the risks and consequences of triggering Earth system tipping points (Winkelmann et al., 2025), (ii) residual warming and global to regional Earth system changes under long-term zero emissions at different GWLs i.e. the zero emission commitment (ZEC) (Jones et al., 2019), (iii) the overall efficacy of negative CO_2 emissions in driving global cooling, including regional patterns of this cooling, and (iv) the reversibility of any induced changes following a negative CO_2 emission pathway.

2. Although the paper is generally clear throughout, my main concern is that the key assumptions underlying how the scenarios are constructed are spread over a very large number of paragraphs. Is there a way to present everything together in an easier form. Possibilities could include an extra "text box", a schematic, or some other display item. It took me quite some time to fully grasp the assumptions, as I had to jump back and forth within the manuscript. Here are the points that should really stand out:

- At the very top level, we set global warming rates we wish our ESMs to follow in the positive emission phase.We cannot invert ESMs, to get out precisely the emissions to track these profiles. Hence, we need to estimate the emissions that get as close as possible to temperature profiles.
- For the upward trajectory, we use the TCRE of each model.
- For the stable part, we assume no net emissions (i.e. net-zero) into the atmosphere (ZEC). And maybe mention this clearly, because for most models, CO2 drawdown balances additional warming towards equilibrium. Hence, why temperatures are stable. ZEC is an emergent property.
- The ramp down assumes a negative value of emissions (i.e. "-X") that were used in the ramp up. Whether GMSAT is reversible on the same timescales as the ramp-up will be an emergent signal.

Good suggestion. Thanks. We have added a Box 1 just before the Fig 1 that reads:

Box 1: Key assumptions and aims of the TIPMIP ESM protocol

- *We aim for a common global mean surface warming rate of 2°C per century across all models in the ramp-up*
- *This is achieved by using each model's TCRE to provide the required CO_2 emission rate in GtC per year.*
- *We approximate warming stabilization by branching runs using net zero CO_2 emissions off the ramp-up runs. The degree of warming stabilization (ZEC) will be an emergent property of each model, likely sensitive to the GWL.*
- *We branch negative CO_2 emission runs 50-years into each zero-emission run, with negative emissions equal and opposite to the ramp-up. Whether the induced cooling mirrors the (ramp-up) warming will also be an emergent property of each model.*
- *We branch a second net zero-emission run off the 4°C ramp-down run at GWL=2°C to investigate potential hysteresis compared to the zero-emission runs branched directly off the ramp-up at GWL=2°C*

3. I realise that the manuscript is mainly a presentation of protocol, but it might be worth adding a little more in the Discussion that [A] the analysis reconfirms with these ESMs that the ZEC is appropriate for Net Zero aims – getting net emissions down to zero does stabilise the climate.

Generally, at GWL=2C this is true (in an ensemble mean sense). It is less true at higher GWLs (e.g. 4C). We have a paper in review on the sensitivity of ZEC to GWL (based on 1 model only) and the underlying causes of this sensitivity. So, we agree with the reviewer, up to a point, and look forward to carrying out a multi-model / multi-GWL analysis of ZEC to confirm (or not) this point. We have added some text where we discuss figure 2, to point this out and highlight the potential sensitivity of ZEC to GWL. This reads as follows:

The ensemble mean ZEC at 2 °C is very close to zero, indicating for GWL=2°C that achieving net zero CO2 emissions may allow global warming to stabilize. This spread in the ZEC response increases at GWL = 4 °C, with an increased tendency for positive ZEC, with a few models suggesting the ZEC response becomes more positive with time through this zero-emission run.

4. And [B], in the Allen et al. (2009) paper on TCRE, I think there was always an assumption that this was well-defined for positive emissions, specifically increasing temperatures. It is of much interest that the reversal also seems to give a good near-linear correspondence between -X and d_Temp / d_time, and of a similar ratio.

In the paper we only show ramp-down results from 1 ESM (GFDL_ESM2). This model does look to be relatively equivalent in ramp-up/ramp-down warming/cooling rates, for mirror image (positive/negative) CO2 emissions. We have only looked at 1 more model in detail with respect to the ramp-down cooling rate (UKESM1.2) and it also looks to show a similar equivalent warming/cooling rate for the same positive/negative emission rate. We are not in a position to generalize this finding yet. We think the degree of linearity (warming to cooling) may depend on the ZEC response of each model and thus may also be a function of the GWL level a ramp-down starts from. This will be determined through subsequent analysis of the multi-model results. We have added some text in the paper explaining this.

The two negative emission runs branched from the zero-emission runs at GWL=2 and 4°C both have a near symmetrical global cooling compared to warming in the ramp-up, suggesting, for this model, that TCRE is reversible across equivalent positive and negative emission pathways. The degree to which this finding holds for other models will be assessed once the multi-model ensemble is complete.

5. Small things

The paper can appear overly heavy on abbreviations (for instance, the left-hand column of Table 1). It may be worth reiterating at a few key points in the manuscript that these names are actually those associated with the metadata of the simulations themselves. Hence, their importance. One possibility is to put them in quotes, maybe?

We agree the paper is a bit heavy of abbreviations, though we think you are mainly referring to tables 2 and 3. We have tried to address this, to some extent this is unavoidable given the nature of this paper. Where possible we have tried to reduce this.

6. Please reiterate towards the end that the analysis is highly idealised. That is fine, and avoids more convoluted emissions scenarios (e.g. to match some sort of upturned parabola in global temperature). However, in reality, unless there is a massive global economic collapse, an instantaneous switch from business-as-usual to ZEC is unlikely.

Agreed and we have reiterated the idealized nature of the protocol and the second paragraph of section 6 now reads:

While this experiment protocol is highly idealized (e.g. using only CO_2 emissions as a surrogate for all anthropogenic forcing and assuming instantaneous onset and cessation of these emissions), the resulting ensemble can be used to investigate (i) the risks, consequences, and drivers of abrupt/rapid Earth system change (tipping points) under net zero CO_2 emission, initiated at different global warming levels, (ii) the long-term response of the Earth system to net zero CO_2-emissions, (iii) the behaviour of the Earth system under prescribed net negative CO_2-emissions following a specified (magnitude and duration) of global warming overshoot, (iv) the efficacy of net negative CO_2 emissions in driving global cooling, and (v) the reversibility of Earth system change under a pathway of positive (warming), zero (near stabilization), and negative (cooling) CO_2 emissions. The primary advantage of the protocol is that we can control the rate of global warming and (potentially) cooling across models, using a TCRE-based CO2 emission rate specific to each model. This enables a multi-model intercomparison of the above topics in a near common global warming/cooling space.

7. Given the interest in Net Zero, then perhaps mention more strongly that this is often regarded as corresponding to the ZEC runs. Having these key buzzwords of phrases

will enable more people to find this manuscript, and so appreciate that ESMs are now being configured for Overshoot and/or periods of Net Zero.

We have added some text into the second paragraph of the introduction to highlight how global warming overshoot and net-zero emission questions can be addressed with these experiments. This reads:

There is increasing interest in global warming overshoot scenarios (i.e. temporary exceedance of a global warming target that is returned to at some later date; Schleussner et al., 2024, Reisinger et al., 2025) and in the long-term response of the Earth system to net zero CO2 emissions at temperatures warmer than today (Allen et al., 2022, Frankhauser et al., 2022). Key outstanding questions include, the likelihood, timing, and magnitude of overshoot, the consequences arising from such an overshoot, the feasibility of cooling the planet back from a peak warming to a (stabilized) target level, the risk of hysteresis once a stabilized climate is achieved, and the long-term response of the Earth system to net-zero emissions at temperatures warmer than today.

Commented [CJ1]: https://doi.org/10.1038/s41586-024-08020-9

Commented [CJ2]: https://doi.org/10.1146/annurev-environ-111523-102029

Commented [CJ3]: https://doi.org/10.1146/annurev-environ-112320-105050

Commented [CJ4]: https://doi.org/10.1038/s41558-021-01245-w

Our protocol lends itself to addressing a number of these questions. Areas of particular interest include (i) the risks and consequences of triggering Earth system tipping points (Winkelmann et al., 2025), (ii) residual warming and global to regional Earth system change under long-term zero emissions at different GWLs i.e. the zero emission commitment (ZEC) (Jones et al., 2019), (iii) the overall efficacy of negative CO_2 emissions in driving global cooling, including regional patterns of this cooling, and (iv) the reversibility of any induced changes following a negative CO_2 emission pathway.

Reviewer 2

1. Jones et al. propose a new Earth system modeling protocol for CMIP7. The protocol includes a set of CO2 emission-driven simulations that reach specified global warming levels (GWLs) with a common global warming rate. When the specified global warming level is reached, CO2 emissions are set to zero and, after 50 years, to the negative of the positive emissions rate and the experiments is continued until a model's global mean temperature drops below the pre-industrial value. The paper justifies the modeling protocol, lays out the science questions the protocol is set to answer and provides a proof of concept with an initial set of ESM model simulations.

   The proposed framework is compelling for its simplicity but also has disadvantages, which stem from the desire to answer a range of different science questions that, ideally, would entail different modeling protocols. While simulated ESM GMSAT trajectories have a common warming rate until specified GWLs are reached, these trajectories can be expected to diverge after this point due to a range of different GMSAT responses to zero emissions (as shown in Fig. 3b). The ramp-down GMSAT trajectories can be expected to diverge even more due to models' different GMSAT responses to net negative CO2 emissions. Therefore, the only factor that is effectively the same across simulated GMSAT trajectories is the initial warming rate and the global warming level reached.

This is true and we have modified the text to make it clearer that (i) we aim to control the rate and amount of warming in the ramp-up run (through an assumed linear relationship between cumulative CO2 emissions and global warming (i.e. TCRE). We recognise this may not hold at higher warming levels and point this out in the paper but to a first approximation this assumption looks valid), (ii) that models then branch into zero CO2 emission runs at common GWLs after the same amount of warming, (iii) that the global warming response under net zero emissions is an emergent property of each model that can also potentially vary as a function of warming level, and (iv) that while we start the ramp-downs at the same time point in each model's zero emission and use the (negative) mirror image of the ramp-up emissions, there is no guarantee that the cooling trajectory in a given model will be a mirror image of the warming trajectory. Points (iii) and (iv) are questions we think the protocol can help understand.

We have added a Box 1 into the paper just before fig 1. This summarizes the main assumptions of the protocol and now also states the caveats on ZEC and negative emission runs as mentioned above. This Box reads as follows:

Box 1: Key assumptions and aims of the TIPMIP ESM protocol

- *We aim for a common global mean surface warming rate of 2°C per century across all models in the ramp-up*
- *This is achieved by using each model's TCRE to provide the required CO_2 emission rate in GtC per year.*
- *We approximate warming stabilization by branching runs using net zero CO_2 emissions off the ramp-up runs. The degree of warming stabilization (ZEC) will be an emergent property of each model, likely sensitive to the GWL.*
- *We branch negative CO_2 emission runs 50-years into each zero-emission run, with negative emissions equal and opposite to the ramp-up. Whether the induced cooling mirrors the (ramp-up) warming will also be an emergent property of each model.*
- *We branch a second net zero-emission run off the 4°C ramp-down run at GWL=2°C to investigate potential hysteresis compared to the zero-emission runs branched directly off the ramp-up at GWL=2°C*

2. The authors justify this approach with different science questions they intend to address: exploring tipping point behavior at different global warming levels, while quantifying the zero emissions commitment at different GWLs and reversibility in response to net negative emissions. However, the tipping point science question would be better answered with a set common GMSAT trajectories (as could be obtained with the AERAMIP protocol), which would allow to clearly explore the risk of tipping as a function of GMSAT.

We agree that the AERA-MIP experiments are extremely useful for the analysis of tipping risks at different stabilized global warming levels (GWL). We have modified the paper to make it clear that the TIPMIP protocol allows us to analyse models under net-zero CO2 emissions initiated at common (and known) GWLs after a common amount of CO2 forced warming in the ramp-up. These can be compared to the AERA-MIP runs at a constant GMSAT. This difference will hopefully allow some useful collaboration and comparison of findings between AERA-MIP and TIPMIP. We fully intend to do this, and some of the coordinators on AERA-MIP are also co-authors on this paper and involved in TIPMIP more generally.

An important factor in the TIPMIP ESM protocol is its simplicity (though we acknowledge in the paper that this simplicity comes with caveats). Because it is such a simple protocol, it is easy for modelling groups to run, evidenced by 11 ESMs having almost completed the protocol at this time.

3. The ZEC and reversibility questions, on the other hand, would be better explored in a framework with common emission trajectories that entail the same rate of CO2 emissions, the same level of cumulative emissions and the same rate of net negative emissions (as in flat10mip; Sanderson et al., 2025). With the current design, it is challenging to attribute differences in GMSAT response to model differences as opposed to differences in emissions trajectory. It would be helpful if the authors could discuss why they chose not to use the AEAMIP protocol to derive a set of common GMSAT trajectories for the tipping point exploration, acknowledge the challenges of the current modeling protocol for the attribution of differences in results, and discuss how they are planning to isolate the effect of different GMSAT and emission trajectory characteristics on outcomes.

We agree that the flat10MIP protocol is an excellent alternative for calculating TCRE, ZEC, and the response to negative CO2 emissions. The primary difference between the two protocols is that in TIPMIP we control the rate of warming across models (in the ramp-up phase), and the GWL level at which zero emissions are initiated. In flat10MIP, commonality across models is instead imposed with respect to CO2 emissions. We expect a number of ESMs that have run the TIPMIP protocol will also run flat10MIP. Joint analysis of the responses in common-GMSAT vs common-CO2-emission space will hopefully help in understanding some of the differences seen between the 2 protocols when run with the same model. Some of the flat10MIP coordinators are also co-authors of this paper and active in TIPMIP, thus we envisage a productive collaboration between the two MIPs.

We have made it clearer in the paper that the TIPMIP protocol does not guarantee warming stabilization at a given GWL, rather it is an investigation of the response to net-zero emissions initiated in models at the same GWL, after the same amount and rate of warming, and then also allows an analysis of the efficacy of net-negative emissions after a (near) common pathway of warming ramp-up and (50 years) of zero emissions have been simulated in each model. As mentioned above the AERA-MIP protocol is definitely very useful for studying the Earth system under a stabilized GMSAT. The primary advantage of the TIPMIP protocol lies in its simplicity, the focus on net-zero, and the aim to constrain models to similar warming and cooling rates.

4. The paper is generally well written but some sections feel lengthy and somewhat repetitive and could be streamlined. Also, citations often seem to be limited to the authors' own work and should include a wider body of literature on TCRE, ZEC and reversibility (see specific suggestions below).

We have gone through the paper and tried to reduce its length and avoid repetition. We have also added in the suggested extra references on TCRE and ZEC, which we thank the reviewer for pointing out.

5. Specific comments:

(numbers refer to line numbers)
76: 'net' negative CO2 emissions (here and elsewhere).
Done

92: It would be helpful if the Introduction could include an overview of the MIP landscape to provide context for the proposed modeling protocol. You do this in section 3.4 but part of it should be in the Introduction as well.
We have added a paragraph at the end of the Introduction section briefly mentioning the CMIP7 landscape as it stands now, highlighting the MIPs TIPMIP has some commonality with. This reads as follows:

TIPMIP ESM experiments began as part of CMIP6plus (an extension to CMIP6 after the formal project had closed). This paper describes the experiment protocol in detail as preparation for participation in CMIP7 (Dunne et al., 2025). 38 MIPs have so far registered to take part in CMIP7, with the full list available at https://wcrp-cmip.org/mips/#registered-mips. TIPMIP has some commonality, and potential for collaboration with a number of CMIP7 MIPs, including C4MIP, CDRMIP, AERA-MIP (Silvy et al. 2024), as well as ScenarioMIP (van vuuren et al. 2025) and the flat10MIP experiments (Sanderson et al., 2024) that are part of the CMIP7 Assessment Fast Track (Dunne et al., 2025). We discuss areas of potential collaboration with these MIPs in section 3.4

Commented [CJ5]: https://doi.org/10.5194/gmd-18-6671-2025

Commented [CJ6]: https://doi.org/10.5194/egusphere-2024-3765, 2025

165: Include reference to Matthews et al, 2009 (first paper to introduce TCRE).
Done

166: 'equation 21': describe or reproduce equation in the text.
Done and reads as follows:

We suggest the method of Arora et al. (2020, their equation 21, reproduced below) to calculate model TCRE values,

$$TCRE = \frac{\Delta T_{2xCO2}}{\bar{E}_{2xCO2}} \qquad (1)$$

Where ΔT_{2xCO2} is the Transient Climate Response (TCR), defined (for a given model) as the change in GMSAT relative to its pre-industrial value at the time atmospheric CO_2 concentrations in a 1pctCO2 run (Eyring et al. 2016) have doubled, occurring after 70 years. $\bar{E}_{2xCO2}$ is the cumulative carbon emissions at the time of CO_2 doubling in the same experiment. TCRE is expressed in units of °C per exagram (1000 Pg) of carbon.

185: what is year 'A'?
Is now defined and only used once

187-188: there will still be a strong transient after 50 years of zero emissions, which will influence the downward GMSAT trajectory.
If you mean that internal variability at year 50 in the zero-emission run will impact the trajectory of the subsequent ramp-down run, then we agree and the route to address this is through more ensemble members. We have run 3 members with a number of

models and can assess this, though not for this paper. If you mean a transient (forced) signal, this will depend on the model and the GWL. For GMSAT, at GWL=2K most models are fairly stable, at GWL=4K less so. Clearly, there will be regional differences below this level of detail. Such analysis is deferred to later studies.

188-189: the ramp down GMSAT trajectory will likely not mirror the ramp up trajectory.
This may be the case and is one thing we want to explore with these runs. The one model shown (GFDL_ESM2M) does look to have a similar rate of GMSAT cooling as in its ramp-up phase. This may not be true for all models and may be related to what their respective ZEC response is at different GWLs. We make it clear in the paper that both the ZEC responses and the GMSAT response in the ramp-down runs are emergent properties of each model and GWL and while we might expect some commonality between the ramp-up warming and ramp-down cooling, this is not guaranteed.

We have added a Box 1 just before Fig 1 where we summarize the key assumptions of the experiment protocol and also point out where warming/cooling responses are emergent properties of each model. The text in Box 1 is quoted above in response to an earlier comment.

364: except that GMSAT trajectories under the TIPMIP protocol will not stabilize exactly.
Agreed. We changed the wording in this sentence to the following:

TIPMIP simulations can be compared with AERA-MIP results to explore, for example, the impact of non-CO_2 radiative forcing on the spatial patterns of climate change in the AERA-MIP runs stabilized at GWL=2°C and the TIPMIP net zero-emission runs started at GWL = 2 °C. A difference between the two protocols is the commonality of model warming rates in the TIPMIP ramp-up runs, as well as the commonality of GWL at which net-zero CO2 emission runs start.

366: Keller et al.: clarify that reference is for CMIP6 CDR-MIP.
Done

379: why 'risk'?
Wording has been changed to:
This may result in only a small number of tipping events occurring in our Tier 1 ensemble

420: Section 4 feels long and would benefit form more concise writing
We have tried to reduce the length.

433: Refer to 2025 tipping point report.
Done

635: meaning of 'connectivity' unclear.
We have changed the wording here to:
Another important question relates to how changes in one location (or component) of the Earth system can cause changes elsewhere.

658: 'net' negative CO2 emissions
Done

691: Include additional references on sink reversal under net negative emissions e.g. Tokarska et al (2015) https://iopscience.iop.org/article/10.1088/1748-9326/10/9/094013
Done

695: This paragraph does not seem to fit here.
Agreed we have removed the first part of this paragraph and merged the remainder with the paragraph above, which now reads:

Carbon dynamics under negative emissions are qualitatively understood, although models differ in the quantitative magnitude of their response. Over the last century, CO_2 emissions have caused a rapid rise in atmospheric CO_2 concentration and natural carbon sinks – both land and ocean – while removing some of this, have not kept up and so increase their sink rates in step with the emissions (Raupach et al., 2014). When CO_2 emissions stop increasing, atmospheric CO_2 declines and natural sinks weaken, eventually saturating, in particular the land biosphere (Silvy et al., 2024). Under negative CO_2 emissions, atmospheric CO_2 concentration strongly declines, and the natural sinks eventually reverse as both ocean and land ecosystems outgas excess carbon dioxide (Tokarska et al., (2015), Jones et al., 2016; Koven et al., 2022). Thus, in the long term, the same processes which lead to natural sinks opposing positive emissions also lead to the natural carbon cycle opposing CO_2 removal, i.e. for every ton of CO_2 removed (e.g. by CDR), atmospheric CO_2 concentrations will decline by less than one ton. Uncertainty in the carbon cycle response to negative emissions is likely dominated by land ecosystems in the same way as it is for the response to rising emissions (Jones and Friedlingstein 2020), although on longer timescales the land sink will saturate more quickly leaving a larger role for ocean uptake (Randerson et al., 2015). The TIPMIP ESM experiments will allow an improved quantification of the various contributions to uncertainty in the carbon cycle response to negative emissions based on the latest generation of ESMs.

705: Move paragraph to section 4.2?
Agreed and we have moved the paragraph and slightly modified the wording in the first two paragraphs of section 4.2