egusphere-2025-3604

The TIPMIP Earth system model experiment protocol: phase 1

Referee Comments & Author Responses – April 2026

Reviewer 1

Suggestions for revision or reasons for rejection

(visible to the public if the article is accepted and published)

Thank you for inviting me to re-review the paper: “The TIPMIP Earth system experiment protocol: phase 1”. I remain convinced this is an important manuscript, and especially given the combination of currently high GHG emissions in tandem with a desire to stabilise global warming at 1.5 or 2.0oC. These two factors imply a high likelihood of a period of global temperature overshoot.

It is during that overshoot period that climate components could potentially tip (depending on their level of inertia), and ESMs remain a key tool for answering that question. In that context, much of the 1.5oC tipping point report has now appeared as a peer-review paper titled: “The implications of overshooting 1.5oC on Earth system tipping elements – a review” – which feels like it should be cited. That paper (https://doi.org/10.1088/1748-9326/ae3cad) explicitly states that more research into the effect of overshooting on tipping is required. This paper should be cited, because it supports the TIPMIP work presented here.

Author: Thanks for highlighting this paper. We have cited it at the end of section 2.

Overall, the authors have responded very well to all comments and requests, making this re-review easy. The new paragraph on the rationale for investigating overshoot is appreciated, and the new display item, Box 1, showing the TIPMIP protocol, should make it easier for readers to fully understand the approach. Thanks for creating that.

The query about ZEC implying that Net Zero may be achieved by causing a subsequent constant temperature remains of much interest. Hence, the extra text the authors have added of noting this might not be applicable at higher temperatures is really important. That is, text “This spread in the ZEC....with few models suggesting the ZEC response becomes more positive with time”. It might be appropriate to re-iterate this important point in the Discussion part of the manuscript.

Author: We agree that the sensitivity of ZEC to GWL and duration of zero emissions at a given GWL are both important. These two points are being actively investigated using the TIPMIP ensemble and will be reported on soon in a new publication. We have reiterated the importance of these points just after figure 2.

The extra paragraph noting that the experimental format is highly idealised is important. Thanks. This does not detract from the paper and will encourage others to discuss with policymakers what less-idealised (but non-SSP) profiles might look like.

Author: Agreed.

At this final stage, I would recommend that the authors make one final sweep through to check for any ambiguities, grammar issues, or typos.

Author: This has been done.

The format of the diagrams could be improved to bring them up to publication standard. For instance, in Figure 2, make the bounding box less bold, remove the dotted grid, and place the "y" labels (ppm and °C) next to the numbers, not as annotations within the panels. The legend "Global warming levels" looks wrong – these are "global warming thresholds"? Similar issues for the other figures.

Author: Thanks. We have made these changes in the new version.

Otherwise, I think the paper is broadly ready for publication.

Reviewer 2

Suggestions for revision or reasons for rejection

(visible to the public if the article is accepted and published)

I have reviewed the manuscript by Jones et al. for a second time. I appreciate the authors' efforts in revising the manuscript and addressing a number of my previous comments. However, I found that the revisions in response to several key points remain incomplete. As a result, further revisions are needed before the manuscript can be considered for publication. Below, I summarize the main outstanding issues, many of which reiterate points raised in my initial review. Additional detail is provided in the annotated manuscript.

• The introduction would benefit from a clearer justification for why a new MIP is needed in addition to existing CMIP7 MIPs. I suggest expanding the final paragraph of the introduction to more explicitly describe what distinguishes the proposed MIP from others.

Author: This has been done. See our response below to the comment at L109

• As noted in my earlier review, a limitation of the proposed protocol is that it does not control for cumulative CO_2 emissions, which are the primary driver of temperature responses under zero and net negative emissions. This limitation should be clearly acknowledged and discussed.

Author: We have hopefully addressed this sufficiently in the revised version.

• The manuscript continues to include statements suggesting that zero emissions result in stable temperatures. I recommend reviewing the text carefully and revising such statements for accuracy (several examples are flagged in the annotated manuscript). In Section 4, the newly introduced language relating net zero CO_2 emissions to global warming levels is difficult to follow and would benefit from clarification.

Author: We have reworded many of the references to warming stabilization and make it clear that the zero-emission runs **start** at common GWLs, with no guarantee they will end at the same GWL.

• In several instances, the manuscript refers to "TCRE reversibility." As TCRE is a metric, it cannot itself be "reversed," in contrast to changes in the Earth system. I suggest revising this terminology to refer either to the reversibility of global warming or to TCRE "symmetry," i.e., the extent to which TCRE under net negative emissions mirrors TCRE under positive emissions.

Author: This has been changed following the reviewer's suggestion

Additional: pdf uploaded by reviewer with specific comments

Author: Line numbers below are with reference to the annotated comments in the reviewer's PDF. We respond to each comment individually with TRACK left on in WORD to highlight where the text has been updated to address these comments. Italics below indicates new text taken from the updated manuscript.

L63: Added *"and potentially also the rate of cooling"*

L67: Changed text to *"under a pathway of positive (warming), zero, and negative (cooling) CO_2 emissions."*

L83: I would view this as an important science and policy question (i.e. the timing of a warming overshoot for a given emission pathway depends on the Earth's (equilibrium and transient) climate sensitivity. The likelihood of overshoot is also dependent on this.

L84: text changed from risk to "*potential for*"

L96: Text changes to "*net negative*"

L109: We have added the following text outlining what we view as the important aspects of TIPMIP. Some of these are common to other MIPs but (to the best of our knowledge) only in TIPMIP are all of these true.

> *An important feature of TIPMIP is that global warming in the positive emission (ramp-up) phase are common across models, Models branch into net zero emission runs after a common duration and rate of global warming, at the same GWL. Models run with zero emissions for a common time period (50 years) before switching to a negative emission (ramp-down) phase, with all models using the (negative) mirror image of the (ramp-up) positive emissions. Models are required to run in CO2-emission mode, allowing us to assess the co-evolution of atmospheric CO2 and climate under net zero emissions at common starting GWLs. The protocol also allows an assessment of ZEC at different GWLs, over 300 (potentially 500) years. In addition, the protocol enables analysis of the efficacy of negative emissions to drive cooling, from different GWLs, when both the natural carbon cycle and climate are able to freely respond to the imposed changes in atmospheric CO2 (the prescribed negative emissions). The protocol is easily extendable to; (i) sample net zero emission responses at different GWLs (e.g. 1.5°C or 3°C) started from existing ramp-up and ramp-down runs, (ii) sample the change in the rate of cooling and climate reversibility as a function of the prescribed negative emissions (e.g. ½ or ¼*

of the original rate), and (iii) the implications of a delay in the onset of negative emissions (global cooling) from a given net-zero (GWL) run. For example, starting negative emissions after 200 years of net zero emissions rather than 50 years. Many of the simulations needed to make these additional runs will have been completed in the Tier 1 set of experiments and the technical implementation of these extensions is trivial.

L187: Text changed to: "*The most accurate method to calculate TCRE*"

L188: Text changed to: *"Because using the TCR approach may lead to slight deviations from the intended warming rate, this approach can be slightly modified (i.e. TCRE values slightly adjusted) to ensure a 2°C per century warming rate."*

Box 1: Last point changed to:

- *We branch a second net zero-emission run off the 4 °C ramp-down run at GWL = 2 °C to compare a zero-emission run after overshoot to one branched directly off the ramp-up at the same GWL.*

L250: A comment to this effect has been added to the paper addressing the change at L188 (see above)

L268: Text modified to:

The realized cooling rate in each model will depend on the amount of excess heat sequestered during the ramp-up and zero-emission phases, as well as how the carbon cycle, atmospheric CO2 concentrations, and climate, respond in each model during the negative emission phase. We might expect a near mirror image ramp-down global cooling, but this too will be an emergent property of each model, likely related to its ZEC (Koven et al., 2023).

L271: We have added the following text to reiterate this point.

It is important to reiterate our experiment protocol aims to control the evolution of GMSAT across models, acknowledging this may only be approximated in the zero and negative emission phases. In doing this, the cumulative CO2 emissions in each model will differ significantly. For example, while models all branch into zero-emission runs at the same GWL after the same duration of warming, the cumulative emitted CO2 and the atmospheric CO2 concentration at this branch point will differ across models. This is a direct result of controlling for GMSAT at the expense of cumulative CO2 emissions and needs to be considered in analysis of the resulting ensemble.

L392: This has been reiterated in response to the comment at L271 (see text directly above). I think this is sufficient.

L423: CDRMIP: We have referenced CMIP7-CDRMIP plans. There does not yet appear to be a paper on CMIP7-CDRMIP.

L502: Text modified to: *We aim to investigate the risk of tipping events under net zero emissions initiated at different levels of global warming*

L508: Text now modified to*: (i) under net zero emissions started at two different GWLs, (ii) associated with the duration of net zero emissions, and (iii) the potential for reversibility as warming is reduced.*

L564: Text modified to: *global temperatures can be reduced, for example through net negative CO2 emissions.*

L570: Citation included.

L574: Cumulative added.

L575 Sentence deleted and a paragraph added to section 4.3 Reversibility:

> *The TIPMIP experiments allow us to address questions around the potential reversibility of Earth system change. A first question concerns the reversibility of global warming realized in the positive emission (ramp-up) phase. One might expect equal and opposite negative emissions to induce global cooling of a similar rate to the (ramp-up) warming. This is not necessarily the case for the TIPMIP runs, as each model's response during the 50 years of zero emissions will differ, resulting in different GMSAT and atmospheric CO2 values across models at the start of the negative emission run. The spread in these quantities at GWL=4°C may also be greater than at GWL=2°C. Hence, while the linearity of TCRE and the link between cumulative emissions and peak warming suggest TCRE and TCRR are symmetric, for the TIPMIP ensemble this needs to be established*.

L599: Text modified to: *The potential for gradual warming under net-zero emissions may result in system thresholds being exceeded,*

We do not include "cooling". Even though it might occur for zero emission runs at different GWLs, because we are looking at GWLs > present day, a slow cooling is therefore unlikely to cause impacts on ecosystems etc. We added the word "potential" to signify that gradual warming is not seen in all models.

L627: Text modified to: *The TIPMIP zero-emission simulations offer the opportunity to assess the long-term response of the Earth system to net zero CO2 emissions at different GWLs.*

L635 - 650: Bullet points have been reworded. Please see new submission.

L676: Text modified to: *Irreversibility or hysteresis is often related to tipping events (van Nes et al., 2016; Armstrong McKay et al., 2022; Lenton et al., 2023), although irreversible change can also occur as a result of incremental warming or climate change.*

L740a: Text modified to: *It remains unclear how symmetrical TCRE is (i.e. the extent to which TCRE under net negative emissions – also referred to as TCRR (Chimuka and Zickfeld 2026) - is equal and opposite to TCRE.)*

L740b: Agreed. This phrase has been removed: *but to first order we expect global temperatures to reverse by approximately the same amount if CO2 is removed at in the same rate as it was added, particularly from lower warming levels*

L770: Fair point, the word "*uncertainty*" is removed as this is hard to attribute in a quantitative manner to different processes. Text changed to:

> *The TIPMIP ESM experiments will enable quantification of the various contributions to the total carbon cycle response to negative emissions simulated in the latest generation of ESMs.*

L811: Text modified to: *The two negative emission runs branched from the zero-emission runs at GWL = 2 °C and 4 °C both show near symmetrical global cooling compared to warming in the* *ramp-up. The degree to which these findings hold for the multi-model ensemble will be assessed in forthcoming studies.*

L842 : We have added a table showing the annual emission rate applied to each model during the positive emission phase of the protocol.

Reviewer 3

Suggestions for revision or reasons for rejection

(visible to the public if the article is accepted and published)

The manuscript by Jones et al. lays out the TIPMIP modelling protocol. The manuscript is very well written, fully contextualized within the body of the existing literature and other MIPs, and the research design and implementation are convincingly laid out. I commend the effort and am convinced that the TIPMIP experiments will provide important insights in a timely manner for the forthcoming IPCC AR7.

Author: Many thanks for these kind words.

I nevertheless have some overarching comments on the modelling protocol. I think some comparably minor amendments can greatly enhance the policy relevance of the experiments. Given the advanced state of the Tier 1 model runs and the time pressure for AR7 it might be difficult to implement them. Yet I hope the authors see value in the feedback and will explore options on how to consider them:

1. The ramp up/down symmetry. Other experimental protocols, i.e. CDRMIP, also do this. Yet, I've not seen a real scientific justification on why it should be implemented like this (other than testing TCRE/TCRR but this doesn't require forcing symmetry). There's ample evidence implying that the ramp up/down phase in the real world will not be symmetric in terms of emissions. Removing around 40 Gt CO2 per year is well outside the range implied by IAMs (AR6 overshoot scenarios would cluster more around 10 Gt removal by the end of the century, and there are already considerable feasibility and sustainability concerns with that). Which means the ramp down segments of the protocol suggested here are skewed optimistic to an unrealistic level. This matters for the interpretation of the results. And there would be a simple fix, i.e. just halving the rate on the way down (still optimistic, but a big step towards a more 'realistic' overshoot trajectory). Ultimately, for Tier 2 it could also be interesting to explore the implications of different rates of decline.

Author: Agreed.

The reason for symmetrical positive and negative emissions was so we could start by investigating the degree of symmetry in warming and cooling (TCRE and TCRR) for mirror image emissions. We agree the imposed negative emissions are unrealistic. For this reason, we already recommended in section 3.2 (potential extensions to Tier 1) that groups consider running with 50 or 25% of original negative emission rate. We have now also added 12.5% as a suggested rate and promoted this higher up our list of suggested extensions. We have added some lines explaining our rationale and that the imposed negative emission rates are unrealistic in section 3.1, as follows:

> *"We choose a negative emission rate in the ramp-down that is a mirror image of the ramp-up to investigate the degree of symmetry in global warming and cooling under opposite and equivalent positive and negative emissions. This level of negative emissions (ranging from -8 to -18GtCyr-1 depending on each model's TCRE) is clearly far greater than anything technically feasible today and probably also in the future, Gidden et al., 2025). We therefore recommend in the potential extensions to the phase 1 experiments (section 3.2) that groups repeat the ramp-down phase using smaller negative emission rates (e.g. 50, 25, 12.5% of their original rate)."*

When we developed the protocol (using the UKESM model) we began with +XGtC/yr as the positive emission rate and -X/2 as the negative rate. When this was discussed with other ESM teams the majority view was to start with -X for the ramp-down and then follow it with -X/2, -X/4. Some models have run -X/2 and we hope others will do this also.

2. I'd like to urge the authors to consider to add a 1.5°C GWL to the Tier 1 experiments, or as a Tier1b setup. I appreciate the beauty in symmetry in the protocol (0,2,4°C). But as a result, the protocol misses important policy benchmarks. In fact, to inform a current discourse, having a 3°C upper limit (close to current policy) arguably would have been more targeted. It seems that ship has sailed. But more importantly, I think a 1,5°C GWL should be added, including a 1.5°C stabilization run and a branching off from the existing 2°C ramp down at 1.5°C. This comes with limited extra effort given the protocol already in place, but would greatly enhance the policy relevance of the insights gained as well as increase the utility of the simulations i.e. for analyzing overshoot pathways with a 1.5°C reference point including in the context of the AR7.

Author: This was extensively discussed when we developed the protocol and we fully appreciate the policy interest in, for example, a 1.5K zero-emission run branched off a ramp-down from GWL=2K. In the end the group decision was to start with 2K and 4K for our target GWLs; (i) to initially keep the number of runs small, (ii) that warming to 2K is (sadly) fairly likely, (iii) 4K was chosen as a "high scenario" level to reach and also a GWL where tipping events might occur more readily (though this is still to be established). A 1.5K zero-emission run branched off the 2 ramp-downs already in the protocol, and from the ramp-up, are recommended in section 3.2 (potential extensions to the Tier 1 protocol) and we have reiterated the importance of these runs. Some models have already made these, and we hope others will soon. There are already plans to compare the transient climates at GWL=1.5K in the positive emission (ramp-up) and the 2 negative emission (ramp-downs) and to contrast these findings with GWL=2K and 3K in the same runs. This will be a start towards considering reversibility and overshoot related questions with 1.5K as the target GWL.

3. As a general consideration for the discussion of results, the question of what plausible reversibility scenarios are needs to be reflected. Recently, Gidden et al. (2025) suggested a prudent limit for underground CO2 storage equivalent to 0.4-0.7°C of GMST reduction. Yes, there's marine CDR. But reversing 4°C of warming is far outside this range. Also reversing 2°C is questionable. Another argument to include a 1.5°C stabilization level.

Author: We have added some text in section 3.1 (see response above in point 1) pointing out the implausible (idealized) negative emission rates used and cite Gidden et al., (2025). We also encourage groups to run with 25, 12.5% of their original values in the potential extensions section.

Minor comments:

L230: I agree simplicity is a sometimes a plus. But I'm not sure I'd claim it's the "main positive aspect".

Author: Modified. Being simple encourages lots of groups to run the protocol.

L288: This problem goes well beyond, because the amount of CO_2 emitted is model dependent. Which means that atmospheric CO_2 concentrations and marine biogeochemistry outcomes will differ across the model ensemble. It might be useful to discuss the implications of this in a bit more detail. Because it also means that i.e. effects related to atmospheric CO_2 concentrations will differ across the model ensemble – which are relevant i.e. for Amazon tipping.

Author: We expanded a bit on this, stressing the difference in cumulative and atmospheric CO_2. The new wording is:

> *As a result, at common GWLs, models will have different atmospheric CO_2 concentrations and will have absorbed different amounts of emitted carbon into the ocean and terrestrial biosphere, impacting responses such as plant photosynthesis or ocean acidification. These caveats should be considered when analysing ecosystem responses.*

L320: I think adding 1.5°C and 1°C as landing zones is crucial. I think 1.5°C should be elevated to a Tier 1, even. This would greatly increase the policy relevance of the protocol as well as the utility in the context of analysis of overshoot pathway outcomes beyond tipping points. For interdisciplinary overshoot pathways, the space between 1.5°C and 2.5/3°C is of particular interest. Less so 4°C.

Author: See our response above under point 2. We discussed both 1.5C and 3C in developing the protocol and the benefits of including them. Our initial aim was for a small set of easy-to-run experiments that were useful for looking at long-term tipping point risks and reversibility, sampling a number of GWLs. 2K was considered both reasonable for a signal and in the policy interest realm. 4K was agreed as (i) a high (but not implausible) overshoot and (ii) a GWL where tipping risks may be increased. Also, the degree of symmetry was considered a plus for such idealized runs. In designing these, we took into account that running 3K and 1.5K zero-emission runs would be relatively straightforward as the required ramp-up and ramp-down runs would have already been done. Hence, they are recommended as Tier 1 extensions. Within the EU project OptimESM, 4 ESMs have already run the 3K zero-emission run off the ramp-up and a few groups have also run 1.5K zero emission runs off the ramp-up and ramp-downs. We agree 1.5K is very relevant for policy but prefer not to change the Tier 1 runs now (after so much discussion with modelling groups) but will encourage groups to expedite 1.5K zero emission runs as soon as possible.

L499: Indeed, it sets the objective to 'hold warming well below 2C". Which makes a 2C stabilization a questionable choice for policy purposes.

Author: Although perhaps a likely outcome for the real-world. But, agreed that a 1.5K zero emission run off a ramp-down from a 2K zero-emission run is a high priority and we are encouraging groups to make these runs.

L504: In most ESM simulations, also AMOC is recovering under stabilization at least at lower levels of warming. Introducing a number of changes.

Author: This is true and is something that is being considered in analysis of these runs. This change can impact both the degree of warming stabilization and the atmospheric CO2 in the zero emission runs. To what extent a recovering AMOC is correlated with ZEC responses across models and GWLs is something we hope someone in TIPMIP will look into.

L498: The rationale presented in this section has a considerable blind spot. It fully applied to stabilization at 2°C. But 4°C? This is well above the current policy trajectory. Of course it could still happen, but then probably because of a stronger climate response. Links to ISIMIP and other communities are being made here. But why should they run an unrealistically high scenario, opening the climate community up once again to scaremongering criticism. I'd fully support a 3°C stabilization. But find it difficult to follow the argument of why 4°C stabilization is a good choice other than for understanding extreme earth system responses.

Author: ISIMIP connections are mainly around the ramp-up, 2K net-zero emission run, and ramp-down and we have explained to this community that these are idealized runs. One topic we hope to be able to address is the "societal impacts" of a tipping point being triggered (say at GWL=2K) and compare the impacts of the tip against the societal impacts of incremental warming to GWL=2K (alone) without the tipping event. i.e. what are the impacts of a given tip (of, say, AMOC, Amazon) and how are these conditioned by the warming level. This is a somewhat different target to the standard impacts of a warming climate that ISIMIP address.

As mentioned above, in the EU OptimESM project 4 ESMs have already run the 3K stabilization and some of the European ISIMIP groups will use this data. Our main aim for TIPMIP ESM was (i) a small number of simulations, (ii) that are useful to look at tipping risks at different GWLs (iii) that allow reversibility questions to be looked into, and (iv) that can be used to force offline models. This latter set of modelling groups were keen to go for a warmer net-zero run in the expectation tipping events might more frequent.

It is unlikely that ISIMIP would use the zero-emission run at 4K.

L546ff: Some key questions are missing here.
What are the benefits of pursuing long-term temperature decline compared to stabilization at higher levels?
Can long-term critical thresholds of tipping elements be temporarily exceeded without triggering irreversible dynamics? Differently put: can pursuing an overshoot pathway pull us back from the brink?

Author: Thanks. We have added these points to the paper,

L655ff: The preliminary results presented in this section are already very interesting. A few suggestions: Consider plotting ZEC (50,100, longer?) relative to the realized warming. Of course, only two data points but could still be interesting. With Tier 2 experiments down the road this could allow for a more quantitative statement on ZEC dependency on peak warming.

Author: We deliberately kept the results section to a minimum and just provided a few examples to show how the protocol works. The primary aim of this paper is to describe the protocol in sufficient detail so that (i) modelling groups can easily run the experiments and (ii) potential users of the data (e.g. climate analysts) understand how the data was produced and the caveats to the protocol. Numerous analysis activities are now starting using the first set of multi-model results, including one specifically looking at ZEC in the multi-model ensemble, considering different GWLs and duration of zero-emission. This particular paper has progressed a fair bit and I imagine it will be submitted in the autumn of 2026. I would prefer not to add more analysis to this paper and rather defer results to the more detailed analysis work that has started across TIPMIP groups.

L685: It's nice to see that there are some first results here. But am not so convinced of the utility of showing single model results. I suspect that for the final manuscript you'll have more models. Else I'd suggest reconsidering the inclusion of this figure and the related discussion in the manuscript.

Author: We included this one model just for illustrative purposes to show how global mean surface air temperature and atmospheric CO_2 evolve over the full simulation pathway. At the start of section 5 we have explained that we show this solely for illustrative purposes and robust conclusions can only be drawn from subsequent analysis of the multi-model ensemble. This analysis has started.

L699: I'd caution to draw this conclusion. Yes, the signal is small compared to 4°C, but maybe not so small in relative terms. A warming/cooling of 0.2°C is 10% of the total signal, or 440 Gt CO_2 emission/removals. The implications for mitigation pathway design would be far reaching (44 years of removals...).

Author: We have modified the text in this section to highlight the absolute and relative uncertainties in ZEC and also indicate that this is being actively looked at and will reported in a companion paper soon.

New text reads:

> *The ZEC response (trend in GMSAT) at 2°C varies across models. A few models show negative ZEC and a few positive. While the ensemble mean ZEC is close to zero, it is important to highlight the uncertainty in ZEC at 100 years is ~+/- 0.2°C (or +/-10% of the total warming). The spread in ZEC increases at GWL = 4 °C, with an increased tendency for positive ZEC across models, with a few models also suggesting ZEC becomes progressively more positive with time through the zero-emission run. The sensitivity of ZEC to the level of global warming at which zero emission are initiated, as well as to the duration of zero-emissions, is being investigated using the TIPMIP multi-model ensemble and will be reported in a forthcoming paper.*