Review of “Asymmetry in carbon cycle feedbacks and transient climate response under positive and negative CO_2 emissions” by Chimuka and Zickfeld.

This is a clear and well written manuscript that explores the symmetry of response of the carbon cycle under positive or negative emissions of CO_2. This is an important question, and one of increasing interest as both science and policy are required to think more about aspects of overshoot and reversibility. Much more is known about the climate system in a steadily warming world than one where CO_2 removal leads to cooling. This manuscript contributes to a research agenda to better address this issue.

The paper uses the UVic earth system model of intermediate complexity. This is a model widely used in the literature and suited to exploring this question. While the results from this model are interesting, and the analysis sheds light on processes involved, I find the value of the paper is more in terms of testing an experimental design which could be adopted more widely by other models. We know that such responses are model-dependent and even the sign of behaviour of terms such as ZEC can vary from model to model. So it is not so much that these results are definitive, but that they lead the way for other models to follow and for a wider research initiative into symmetry and reversibility.

We thank the reviewer for the positive and constructive comments.

I have some minor comments which I list below which I hope the authors find useful to clarify some points.

I do, though, have a major comment about the design and intent of the paper. The more I think about this question, the more I realise that there is a difference between “symmetry” and “reversibility”. It may feel nuanced, but I believe it is important.

The difference is that “symmetry” - as defined in this study - is asking the question “do we see the same response if we go upwards from a starting point or downwards from that same point?”, whereas “reversibility” would ask “do we see the same response going up from one point to another as we do going back down from the second point back to the first?”

Previous studies of reversibility (e.g. Boucher et al 2012, or CDRMIP simulations) have addressed the question of “going back down again”, BUT they have been contaminated by the problem that they have very large legacy effects on the way down because the simulation follows immediately from very strong upwards forcing and is not in steady state. The current authors know this well and discuss it in detail in Chimuka et al (2023). The present study therefore represents a novel experiment design and analysis by equilibrating at a higher level (here 2xCO2), thus avoiding the legacy response in the negative emissions phase. I think this is vital and a very nice experimental design.

However, if the question of interest is whether or not we can recover a previous climate state by the same path in reverse (i.e. does TCRR = TCRE?) then I am not sure that the current experiment answers exactly that question. The asymmetry found in this paper may be because the system response differs above 2xCO2 from below. For example, if TCRE is state-dependent (e.g. lets assume the TCRE gradient gets less steep at higher warming), then we would expect TCRE measured above 2xCO2 to be lower than TCRE measured up to 2xCO2. This would manifest as an asymmetry about 2xCO2 – that's true. But it does not necessarily imply that reversibility is along different lines. TCRR and TCRE could still be identical to each other. Does that make sense?

A more direct answer to the reversibility question would be to compare your ramp-down experiments here with the ramp-up from 1xCO2 up to 2xCO2. So you can measure a TCRE going from 1xCO2 up to 2xCO2, and then a TCRR going back down (but from an equilibrated state at 2xCO2). I think this then answers the reversibility question.

I think you must have such experiments, but I realise that requesting them (and new analysis) to be added here is a big task. As the manuscript stands it does indeed answer the stated question regarding "symmetry". But my challenge to the authors is whether or not that is really the target question? Given that work like this will undoubtedly be influential in experimental design for CMIP (CMIP7 and beyond), then making sure that we address the correct questions is as important as making sure we design the experiments to answer them. One option is to keep the present manuscript and be clear that the question asked is on symmetry but does not necessarily reflect reversibility. This opens up the opportunity for a follow-on paper on reversibility. Or a second option would be to add to the current study an analysis of positive vs negative emissions up vs down between the same points.

I appreciate a request for extra analysis is never welcome, but I hope the authors see this a constructive suggestion.

Chris Jones

We thank the reviewer for these important comments on the distinction between symmetry and reversibility. As the reviewer correctly points out, our experimental design is best suited for answering the question on symmetry. By diagnosing positive and negative emissions from the same equilibrium state, we investigate whether the climate impact of a given amount of emissions can be balanced by an equivalent amount of removals.

With that said, we do see the policy relevance of exploring the reversibility question by comparing our ramp-down experiments to a 1% yr^{-1} ramp-up from preindustrial to 2xCO_2 and had already conducted the required model simulations. Therefore, we have included an additional section ***3.4: An Alternative Experimental Design for Quantifying Carbon Cycle Feedback and TCRE Asymmetry*** *based on this experimental design suggested by the reviewer. Overall, we find that the magnitude of carbon cycle feedback asymmetry is much smaller (by an order of magnitude) in this experimental design, except for the ocean, where the magnitude of asymmetry more than doubles. The smaller magnitude of asymmetry arises because temperature* *change is more symmetric and the range of CO_2 concentrations in both trajectories is the same. The sign of the TCRE is consistent with the main experimental design, whereas the sign of the carbon cycle feedback asymmetry is positive for land feedbacks and negative for ocean feedbacks.*

Minor comments/suggestions:

- Have you thought about pitching the experiments in emissions rather than concentration forcing? I have been a proponent of using concentration-driven runs to diagnose the alpha/beta/gamma feedback metrics because they are in units of "per ppm" so prescribing concentration as the boundary condition makes sense. But for metrics like TCRE/TCRR which are "per GtC of emission", then prescribing emissions makes sense as a boundary condition. Then it is relatively easy to prescribe simulations with symmetric positive/negative emissions. The consequence here of imposing symmetric % change in concentration is very unsymmetric emissions rates (your fig 1b).

  This is something we considered in the course of our study. We opted to use concentration-driven simulations so that we could quantify feedbacks using the integrated flux-based feedback framework and maintain consistency with the feedback framework to ensure comparability between the Jones & Friedlingstein and the direct approach to quantify TCRE. However, one thing we neglected to mention in the manuscript is the time points at which we evaluated TCRE/TCRR. Unlike the feedbacks parameters, which were evaluated at 3xCO2 and 1xCO2 (yr 43) for both positive and negative emissions simulations, we chose to compare TCRE vs TCRR at the same level of diagnosed emissions for both the 1% yr^{-1} and flat10 simulations.

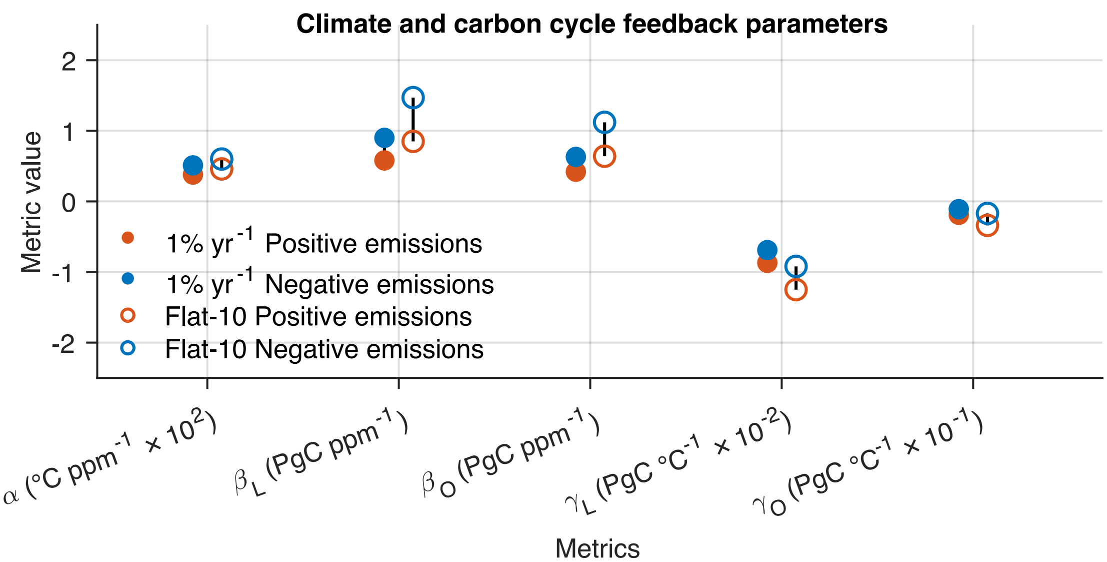
  For example, in the positive emissions simulations, diagnosed cumulative emissions are approximately 800 PgC, whereas cumulative diagnosed negative emissions are -970 PgC in the negative emissions simulation. Therefore, we chose to evaluate TCRE at 800 PgC from the positive emissions simulation, and TCRR at -800 PgC from the negative emissions simulation. In addition, we also evaluated TCRE and TCRR at +/-800 PgC from the flat10 simulations for consistency in the rate-dependence analysis.

- The fact that it only needs 43 years to reach 3xCO2 feels very short – are you comfortable this is long enough for a reliable signal? Potentially should we look at lower rates and longer runs? If this experiment was adopted for example in ESMs with significant internal variability would this be long enough to robustly diagnose T and carbon fluxes?

  We agree that 43 years is a relatively short time period, especially for ESMs. This was partly the motivation for additionally exploring feedback and TCRE asymmetry from the flat10 simulations as those metrics are evaluated after 100 years. Based on our rate-dependence analysis, the sign of the feedback and TCRE is independent of rate of *emissions (and therefore, time of model integration), and the magnitude exhibits little to no dependence. We do, however, agree that simulations with a longer time of integration, such as the flat10 simulations, would be better suited for multi-model analyses.*

- Figure 5 is useful – can you include alpha on here too?

  Figure 5 has now been updated to include alpha. Please see the updated figure below:



- Sec 3.3 on quantifying the asymmetry and using Jones & Friedlingstein framework. I think you could be clearer here that the J&F framework is simply a way to break open TCRE to see which components matter – but yes it might be approximate. The delta-T/CE and delta-T/CR remain the correct answer and the actual definition of TCRE and TCRR (you say "conventional calculation" – I would make this stronger and say "known values" or similar). Just to be clear that J&F isn't a rival definition, just a means of understanding

  Noted. We have changed the wording to "direct calculation".

- Fig 6 – instead of plotting these vs time, I wonder if plotting vs Emissions is more relevant? We know the diagnosed emissions are not symmetric so we wouldn't expect the T vs time to be the same even if TCRE=TCRR. So maybe plotting in emissions space here makes more sense

  We thank the reviewer for their comment. As mentioned in a reply to a previous comment, we plotted TCRE and TCRR against time, but evaluated the metrics at the same level of cumulative diagnosed emissions. We also chose to plot the remaining metrics against time because they track CO_2 concentrations (which are symmetric) rather than cumulative emissions (which are asymmetric).

Citation: https://doi.org/10.5194/egusphere-2025-6405-RC1