

Response to Anonymous Referee 1 for the manuscript

CMIP6 simulations with the compact Earth system model OSCAR v3.1

Yann Quilcaille^{1*}, Thomas Gasser², Philippe Ciais³, Olivier Boucher⁴

¹ International Institute for Applied Systems Analysis (IIASA), 2361 Laxenburg, Austria;

* now at Institute for Atmospheric and Climate Science, Department of Environmental Systems Sciences, ETH Zürich, Zürich, Switzerland

² International Institute for Applied Systems Analysis (IIASA), 2361 Laxenburg, Austria

³ Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, Université Paris-Saclay, CEA – CNRS – UVSQ, 91191 Gif-sur-Yvette, France

⁴ Institut Pierre-Simon Laplace, Sorbonne Université, CNRS 75252 Paris, France

Correspondence to: Yann Quilcaille (yann.quilcaille@env.ethz.ch)

We thank a lot the Anonymous Referee #1 for the comments. We think that their integration of these comments have been important to improve the quality of this manuscript. In the following response, the original answer is in *black italic* while the answer is in *green*.

Here is the summary of the modifications brought to the text:

- Improvement in the structure of the manuscript: reorganized the sections on the diagnostic of the model, moved to the appendix the sections on the behavior of the model
- More discussion in the experimental setup:
 - o Description of OSCAR, with more emphasis brought on descriptions existing in the literature
 - o New conceptual figure for description of the model
 - o Post-processing of OSCAR more detailed
 - o New conceptual figure for description of this framework, discussion of its limits
- More discussion in the sections on the diagnostic of the model
- Edit of the abstract
- More active formulations, proofread
- Correction of the figures to integrate the labelling of panels

General comments

Quilcaille et al. present a summary of a number of CMIP6-style simulations performed with the reduced complexity earth system model OSCARv3.1. They briefly describe the model, how they constrained it to create probabilistic results and then go through the results of their numerous experiments.

The paper is clearly the result of a massive amount of effort, and having this description of OSCAR in the literature is beneficial. My issue with the current presentation is that it lacks any punch. There are so many results, and comparatively so little discussion, that it's hard to know what I'm meant to take away from this apart from, "We ran our model heaps". That's not to say that there aren't really interesting results in the paper, it's just that they're overly hard to find.

We thank the Anonymous Referee 1 for this comment. We agree that the take-home messages of this manuscript are hard to identify.

As remarked, the amount of results for relatively little discussion makes it harder to read. To deal with the lack of discussion of the results, we now discuss into more details the sections concerning the diagnosis of the model.

Although the model is of “reduced complexity”, there are many aspects of the model to cover, and we have been thorough. The take-home messages are indeed difficult to find, that is why we have also rearranged the structure of the manuscript, with results & discussions rearranged in two groups. The first group concerns the diagnosis of the model, and has been kept. The second one gathers results where elements of comparison are missing, but that still inform about the behavior of the model. These sections have been moved to the appendix.

We hope that these two categories of changes would provide a better guidance to the reader on what they are looking for about OSCAR.

I think the paper would benefit greatly from improved focus. Many of the experiments require extensive discussion and further exploration. There doesn't appear to be space for all of them in this paper. For those that cannot be fully explored, I think it is better to save them for future papers where they can be explored appropriately rather than having partial explorations (also because I don't think partial explorations belong in the scientific literature).

We agree with you that there is no space for a full discussion about all the aspects of this reduced complexity Earth system model. It would increase the size of the manuscript that is already more than 50 pages.

You suggest to save them for future papers, but is very unlikely that we will have the time to write papers on that many topics. Besides, it is even more unlikely that these would be scientifically interesting in their own right. Most of these sections that you identify are actually elements illustrating the behavior of OSCAR.

We think that this second group, now in the appendix, is still interesting. Showing how OSCAR behaves in different situations gives a sense of what to expect of the model to the reader. Besides, as far as we know, OSCAR is the only reduced-complexity Earth system model of RCMIP to provide details about the runs it performed for RCMIP. Nevertheless, we are willing to remove this appendix if the Editor considers that these results are unnecessary.

Major concerns

Lack of focus

As discussed above, the paper lacks focus. Results are presented from experiment after experiment, with no space to actually explore their implications or what to make of them. Obvious examples are Sections 7.2, 7.3 and 7.5. However, many of the sections left me wondering, “What is the point?”

As answered above, we have reorganized the manuscript to give a more appropriate structure and highlight the results. We have also given more space to discuss the implications of these results when appropriate.

This is unfortunate, because many of the results are very interesting. For example, it is surprising that the ZEC is much higher for 2000 PgC experiments but overall warming doesn't have the same non-linearity. However, there is no further exploration of this.

This part is now discussed into more details.

Similarly, “the carbon stocks still increase in G6solar, even more than in ssp585 thanks to the lower GSAT and despite lower global precipitation.” Is this what we would expect? Or does this point to a clear limitation of the model if we have less rain but somehow more carbon stocks?

Thanks to the new structure and new discussion there, this part is better explained.

Eight different spin ups are done: how do they compare? What does this tell us about the way we make climate projections and any potential bias in the CMIP-style of doing things.

Although the question that you raise is interesting, we will not compare these spin-ups in this paper. As you noticed, we have already enough results and had not enough discussion, adding more results would not move this balance in the right direction. Furthermore, we consider that reduced-complexity Earth system models are not the proper tool to evaluate such biases, because such models are much less sensitive to the spin up itself than full-fledged Earth system models.

The authors also write, “Our results cannot be compared to the final CDRMIP results yet, for they are unpublished, but they are consistent with those obtained with a model of intermediate complexity (Zickfeld et al., 2021).” However, they have a great tool to evaluate the questions: if they know how much to trust their model (which they should at the end of this evaluation), then they don’t need to wait for the CDRMIP results and could write a great (separate) paper on their results now.

As you suggest, it could do a great paper, and you are right in saying that it should be a separate one: this is not the objective of this manuscript. However, it is highly unlikely that we will have the time to write a paper dedicated to this question. However, we are of course open to collaborations for comparison and analysis, for instance in projects such as CDRMIP or ZECMIP.

I would recommend the authors reconsider which results to present in this paper, which belong in their own paper and which are best left out. This would probably significantly reduce the length of the paper. It would also improve the abstract, which is currently too long and contains too much detail (it could be re-written to

just focus on key points: Emulators are needed, emulators need to be validated, here we examine OSCAR, strengths are X, weaknesses are Y, last sentence could stay as is).

As explained previously, this manuscript had two types of results: those about the diagnosis of the model, with elements of comparison to the literature, and those about the behavior of the model. We changed the structure of the paper accordingly, in a much clearer way, to guide the reader. Besides, we added discussion where it was needed. We think that these modifications would compensate for the aforementioned issues.

We have also edited the abstract, following your suggestions.

Writing

The writing is very slow, i.e. it doesn’t always make clear what the point is. I think this is partly due to a lack of focus as discussed above. It’s partly due to being repetitive (the point about the need for validation is mentioned three times in the first paragraph). However, I think it is also partly due to phrasing. It might help

to swap phrases like, “As illustrated in Table 4, OSCAR v3.1 estimates a ZEC (in the reference case of the esm-1pct-brch-1000PgC experiment) that is within the range of ZECMIP (Macdougall et al., 2020), although the long-term decrease seems to happen later in OSCAR.” with more active formulations (that move all the table and figure references to parentheses) like “OSCAR v3.1 estimates a ZEC (in the reference case of the esm-1pct-brch-1000PgC experiment) that is within the range of ZECMIP (Macdougall et al., 2020), although the long-term decrease seems to happen later in OSCAR (Table 4).” (Yes, I acknowledge the irony of giving advice about punchy writing when my review is probably slightly rambling.) The paper is also in need of a proofread. Reading it this time was overly difficult due to typing and other phrasing errors. This will take some effort, but it will greatly improve the experience for the reader.

We thank a lot the Referee for this comment. We have gone through the paper to work on these aspects.

Vague claims of goodness

The paper has some very vague claims of goodness. Two examples, “It reproduces the responses of complex ESMs, for all aspects of the Earth system.”, and, “the resulting quantitative behaviour of OSCAR remains largely satisfactory”. Both these claims are vague and subjective. I would simply remove them and all others like them from the paper, the reader can judge the quality for themselves based on the results (and likely, the ‘good enough’ level will change depending on the application of interest).

Depending on the context, we have either removed the claim if no points of comparison were possible, or we have given more details for a better comparison. This is in line with the correction for more discussions where needed.