

Response to the Editor 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 want to sincerely thank again the Editor for their approval of this manuscript. Further modifications were brought to the manuscript, following the comments of the referees. 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:

- Improved narrative, focus on the results.
- Removed vague claims of goodness and more quantification.
- New appendix for post-processing.
- New repository for the code of OSCAR for replication.

Thanks you for your revised version of the manuscript. Both reviewers appreciate the improvements on the earlier version. Still, reviewer #2 points to further room for improvement. In particular, I support her/his request to refine the framing and narrative of the paper, and would also ask you to address the other points.

Regarding the in-depth presentation of model behavior, I am fine with keeping it in the appendix. I am looking to a revised version of your paper.

The aforementioned modifications were brought to the manuscript, following Reviewer #2's recommendations. Specifically, the narrative & framing were improved with the following points:

- Although the section on post-treatment of the runs brings useful information for the understanding of the method, it was too long, with the risk to lose the reader. We moved a large part of this section to a new appendix.
- We went through the text to improve it, following the recommendation of Reviewer #2. Several modifications in the text were made to strengthen the narrative. In particular, the introduction has been modified along this line and a new paragraph was added on the recommendation of Reviewer #2 between section 3 and 3.1, serving as reminder.
- Reviewer #2 asked for more quantification. We noticed that the section 3.4 and the abstract were not adequately quantified, contrary to the other sections that benefit from the 5 tables. New additions were made there.

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 want to sincerely thank the Anonymous Referee #1 for its approval on this manuscript. Further modifications were brought to the manuscript, following Anonymous Referee #2's comments. 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:

- Improved narrative, focus on the results.
- Removed vague claims of goodness and more quantification.
- New appendix for post-processing.
- New repository for the code of OSCAR for replication.

Thanks for the careful response to my original review. The improved focus and expanded details on the conditions required for exclusion of ensemble members is appreciated. I'm happy for the paper to be published in its current form.

Thank you very much for your approval. As said before, Anonymous Referee #2 had several recommendations that we followed. In particular, the details on exclusion of ensemble members were moved to a new appendix to strengthen the narrative. We agree that the details that you suggested in the first round are useful for many readers, albeit it may lose some others. This is why we decided to keep them as the first appendix to focus on the main message in the main text.

Response to Anonymous Referee 2 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 want to sincerely thank the Anonymous Referee #2 for the comments. We have carefully read its report and integrated its comments. We deem that it has significantly improved the manuscript, one more time. 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:

- Improved narrative, focus on the results.
- Removed vague claims of goodness and more quantification.
- New appendix for post-processing.
- New repository for the code of OSCAR for replication.

Quilcaille et al. have revised their paper to enhance the focus on model evaluation, leaving pure presentation of model behaviour for the appendix. I think these changes have improved the manuscript. The addition of Figure 1 was also very helpful for me.

Glad to read that it helped.

My major concerns focus on a few key areas, many of which echo earlier comments on the manuscript. I think these can be addressed. I also still think the point of the paper could be made clearer (there were fewer tracked changes than I was expecting to see). Is it not: we have used OSCAR v3.1 in a few places already, here we provide a thorough evaluation of its behaviour over a number of experiments where we have something to compare against, here are the levels of agreement (quantified)? If the paper just stuck to telling this narrative, I think it would be much easier to read.

Thank you for expressing this concern. In the first round of review, the structure has benefited from your comments. However, it is true that the narrative could still be improved. Three main axis have been used to improve the narrative:

- Although the section on post-treatment of the runs brings useful information for the understanding of the method, it was too long, with the risk to lose the reader. We moved a large part of this section to a new appendix.
- We went through the text to improve it, following your recommendation. Several modifications in the text were made to strengthen the narrative. In particular, the introduction has been modified along this line, and a new paragraph was added on your recommendation between section 3 and 3.1, serving as reminder.
- Regarding quantification, we noticed that the section 3.4 and the abstract were not adequately quantified, contrary to the other sections that benefit from the 5 tables. New additions were made there.

On a quick note, there were less tracked changes visible because the 6 pages of text and 8 pages of figure moved from the main text to the supplementary material were not integrated in the tracked changes, to make it easier to read. In this round of revision, all changes were tracked.

I would also note that many of the other reviewer's comments put a pretty high expectation on the authors. In my opinion, many of the questions asked about particular details and choices related to calibration are better explained by the code accompanying the paper (rather than duplicating this information in the paper) or in standalone papers. Adding such things into a pure evaluation paper (whatever that is worth, see comments below) makes it very hard to have focus.

Thank you. We agree that such technical questions would find a more exhaustive description in the code. However, the reasons for the choices made would not necessarily appear in the code. Besides, someone interested in these aspects may get a better understanding of the code by first reading the text and accompanying figures.

To balance the reviews, we moved the Figure 2 and a large part of the section 2.3 to the appendix. We hope that it would bring a better focus to the paper while allowing for readers interested in these details to find the sought information.

Overall, I think the paper now achieves its aim of evaluating the behaviour of OSCARv3.1. However, I do think it could be greatly improved in terms of presentation and clarity.

Thank you. We hope that this new version would have the adequate level of clarity.

The vague claims of goodness persist in this version of the manuscript (even in the abstract). Where they appear, they read like the authors want to be able to say, "OSCAR is good", which is particularly odd, because the authors are very honest about the limitations of their model in many other parts. Again, I would just remove any sentence that uses a subjective judgement, such as 'good' or 'satisfactory'. Just tell the reader what the difference is and they can decide what is good enough based on their own situation.

We acknowledge that the former round of corrections left some of these subjective judgements out. This is now corrected.

The authors have retained their section that focuses purely on behaviour of the model, albeit as an appendix. I can see why they want to keep this section, but I have some further thoughts about this.

The first is this. In the revision process, the authors make statements like the following, "However, we highlight that we would not be able to invest the time to transform these deleted results into future studies, hence they would be lost." The implication is that this is the only chance to publish them. My issue with this statement is that, by saying, "We won't have time", the authors are implicitly saying, "We won't make time". Put another way, the authors are saying that, "These results aren't interesting enough to be worth our time writing up". The problem with this is that it then raises the question, are these results worth anyone's time reading? I think it is ultimately an editorial question whether these pure documentation plots can be included in an appendix or not (they take up space and are disconnected from the main narrative of the manuscript, but you don't have to read them to understand the manuscript so they aren't a negative). However, I still struggle to see why plots of stuff, without any explanation of their implications, belong in the scientific literature (surely they are better captured as part of a tutorial on the model or the model's development repository, where they can be presented without any accompanying narrative?).

When we had written that we would not have the time, it doesn't mean that we consider these results as "not interesting enough to be worth our time writing up". We do see scientific value in these results and foresee appealing papers that would make use of them, particularly on the reversibility of the Earth system using section A.1. The sole reason why we will not be able to make time for such papers is that in modern research, the priorities are set by funding and not by scientific interests. We have to focus on the projects that we are working for, even it means not publishing papers with scientific value. For instance, the main author is now in a different institute, paid on a different project, and the hours spent on this paper cannot be reported on this new project. Then no, it is not because the other papers would not have been interesting, but because we sincerely cannot take this time.

Furthermore, we would not like to have these results completely sacrificed. They would have a good fit in model intercomparison studies, as initially envisioned for the CDRMIP runs. Publishing these results is not enough as a stand-alone paper, but they may be of interest as contributions to other studies. This is why we still want to showcase OSCAR in the supplementary material of this manuscript.

Regarding the "pure documentation plots", I realized that I should have moved the figures of the appendix with their text. I apologize and corrected this mistake. Now, these plots are directly with their corresponding explanations in the appendix. As discussed in the first round of review, these sections on the behavior of the model may bring some insights to the readers interested in qualitative aspects of the model on these experiments, otherwise not shown.

My second thought also follows from a comment by the authors, “We highlight that no other reduced complexity Earth system model has done such a thorough analysis before, and such a paper could be a first step towards better descriptions.” I would agree with this (more or less) and I think it raises a fascinating question about how to document different model versions. The current practice of writing standalone manuscripts is clunky for a number of reasons. Firstly, a complete description of the model is not appropriate for any single manuscript so it never appears anywhere (rather, any user has to piece together the full picture from multiple papers). Secondly, description papers tend to be very long because they have to cover so much territory. Thirdly, they are very hard to write because they don’t have an obvious narrative apart from, “Here is how the model looks/works” (and that narrative isn’t very interesting to most people given models are for insight, not for numbers). Given that current practice is clunky, I would encourage the editors of GMD to give this question further thought: How can model description papers be improved so that they are more useful for authors and readers alike? Are scientific papers even the right forum for such documentation given their focus on narrative and implications? Obviously these questions don’t affect the publication of this paper, but given the authors’ made the comments I thought I would reply.

You are indeed raising a crucial point. This publication is trying to find its place in a “clunky” context, being the description of reduced complexity Earth system models and to a broader sense the description of models. Luckily, reduced complexity Earth system models are simpler than Earth system models, facilitating their full description in a single paper. This is the case for OSCAR, MAGICC, HECTOR and many others. However, it is true that later versions build upon the initial publication, meaning that reading on a model means piecing papers together. Automated wikis are of course possible, although it takes a significant input from a software perspective, and not all modelling teams are capable of that. Versioning tools like GitHub provide a good alternative, albeit less user-friendly as an automated wiki. Therefore OSCAR is on GitHub, although a systematic evaluation of the model’s versions remain to be implemented. This paper is meant to provide a first milestone towards the comparison of future versions of OSCAR. For instance, adding new processes, improving the modeling of the ocean carbon sink, or changing the representation of aerosol chemistry in OSCAR would change its outputs. We consider that this paper would pave the way to a better understanding of what needs to be changed in the model and the effects of the ensuing changes. Of course, there are other solutions to improve model descriptions, each with their pros and cons. We hope that the path that we choose here would give readers a good overview of this model.

The authors refer to the new first section as diagnosis. This language seemed odd to me, I would have used the phrase evaluation because the authors seem to be evaluating the extent to which their model behaves in line with other available literature estimates over a range of experiments. An introductory paragraph at the start of section 3 (before the section 3.1 header) that re-clarifies the point of this section would be helpful (given how long section 2 is).

Thank you, we have edited the manuscript accordingly to your advice.

The paper’s reproducibility would be greatly enhanced if it was clear where an interested party could access the code that sits behind it, particularly the code related to constraining OSCAR. Having the model code available open-source is good, but it isn’t enough to actually reproduce the paper’s results by itself and the descriptions given in the paper are certainly not enough to reproduce the study by themselves.

We have now created a new repository for the code used to run OSCAR, to find the diverging runs, to constrain OSCAR and for the plots produced here. Though, it comes with a warning that this code is very raw, hindering its readability. Again, it is the longer-term plan to integrate evaluation runs with the model’s code on its GitHub repository, based on the work done for this paper, but this will take significant efforts.

Technical corrections:

A selection are listed below, but I would note that the paper is still in need of a good proofread as many of the sentences are still missing words and use odd phrases, which makes reading the paper much harder than it needs to be.

Thank you for such a careful reading of this manuscript, we have integrated these corrections and additional ones after proofreading. Would the reviewer point to any remaining issues, we would be very grateful.