Response to Anonymous Referee 2 for the manuscript

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

Yann Quilcaille[1*], Thomas Gasser[2], Philippe Ciais[3], Olivier Boucher[4]

[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 are grateful to the Anonymous Referee #2 for the comments. After integration of these comments, the quality of this manuscript has improved. 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:
    - Description of OSCAR, with more emphasis brought on descriptions existing in the literature
    - New conceptual figure for description of the model
    - Post-processing of OSCAR more detailed
    - 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

Quilcaille et al describe a large number of simulations conducted with the OSCAR simple climate model, covering some fraction of the CMIP6 experimental design. The paper briefly describes the model and the calibration strategy, and then discusses a number of applied cases.

Clearly a large amount of work has gone into this paper, singlehandledly simulating a large fraction of the CMIP6 experimental design in a single study. However, a side-effect of this is that the paper is somewhat less than rigorous in describing the calibration process and in providing a detailed assessment of the applied experiments in the context of available ESM simulations.
We thank the Anonymous Referee for this comment. We have provided more details on the probabilistic setup. We highlight that this paper is not about the calibration of OSCAR, because this model had already been calibrated before, as described in Gasser et al, 2017 (https://doi.org/10.5194/gmd-10-271-2017).
We have as well given more space to the discussion of the results where needed and changed the structure, to bring more clarity on the take-home messages.

It is also sloppily written in places - with a number of spelling errors and inprecise language. Figures require axis labels and subplot labels throughout.
A thorough proofreading of the manuscript has improved the spelling and language.

Regarding the figures, it is a choice justified by the clarity of the figures. For instance, subplots sharing the same time axis do not need to have “year” and the corresponding tick labels repeated over and over again. A good example is figure 9 for the LUMIP experiments: all columns share the same time axis, then only the bottom row has the axis label written. We think that changing all these figures would add a lot of text and details for information already present, thus decreasing the readability of the figures.
Regarding the labeling of the subplots, we have edited the figures accordingly to the GMD guidelines.

My recommendation is that the scope of the paper is reduced - but greater emphasis is placed on a detailed description of the calibration process, and the senstivities of simple scenarion projections to aspects of that calibration process. Additional experiments (SRM, CDR etc) - are secondary and could be covered in follow-up dedicated studies when the fundamental probabilistic setup is well defined.
We have changed the whole structure for the results, identifying two groups of sections, one about the diagnosis of the model and another about its behavior (SRM, CDR, etc). As you suggest, this second part is indeed secondary. It has been kept as an appendix, because we consider that it is still important to the reader, as a showcase of the behavior of the model. This part illustrates what to expect of the model, which is what some readers are looking for when reading about a model. 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.

Of course, this appendix about the behavior could be removed if the Editor decides that it is not relevant. However, it is highly unlikely that the deleted material would become future papers, because we will not have the time for that many additional studies.

Major issues:
1 - More detail is needed on the basic model structure. This version of the model is not documented in the literature, and the first section needs to give a basic overview of the level of complexity being represented. A conceptual figure illustrating the number of domains, and how they interact, would be appreciated.
We thank the Referee for this comment that has improved the clarity of the manuscript. We have to disagree on the documentation of this version, however. The second sentence of the first paragraph on the description of this version explicitly explains where the version 2.2 is described and where to read about the difference from version 2.2 to version 3.1.

Even though OSCAR is a “reduced complexity” Earth system model, giving a basic overview of what is represented is no easy task. The 4 following paragraphs were actually aiming at that, but your comment shows that it was not enough. For this reason, we introduce a new figure, providing a conceptual overview of the model. Although this is only a conceptual figure, without the details of all equations, datasets and parameters, it provides a good perspective on the level of complexity of the components and their interactions. Of course, we remind the sources for the detailed descriptions of the versions of OSCAR in the caption of the figure.

2 - Perhaps the key aspect of this paper is the effect of observational constraints on projected future climate. However, this is covered quite briefly, and the many degrees of freedom in the calibration process are not comprehensively explored. How do different observations constrain projected warming independently and combined? How are constraints objectively combined? How are prior distributions decided? The paper refers back to Gasser (2017), but the addition of additional constraints in the present paper requires a more detailed description.
We thank the Referee for this comment. The use of observational constraints is important in this paper, but we do not consider this aspect as the main scope of the paper. We highlight that this is constraining, and not calibrating. The calibration concerns the parameters for the equations, which was described in Gasser et al (2017) and Gasser et al (2020). Regarding the constraining, this followed the coordinated effort of RCMIP (phase 2), but going into the details of the (infinite) possibilities of doing such constraining is an entirely different paper in itself. Here, we acknowledge the limits of our approach and the need to investigate this further for future use with OSCAR.

However, what was done for RCMIP is done and will not change, therefore we report in this paper the performance of this precise version of the model.

The manuscript refers to Gasser et al (2017) for the model description, the parameters and the probabilistic framework. However, it is Gasser et al (2020) that provided information on the update of the model and the use of observational constraints.

Nevertheless, we have added details on the limit on how the observational constraints have been chosen and used. We hope that the addition of the new figure and of more details on the post-processing will bring the sought clarifications.

3 - The language is often far too vague on key details, like exactly how CMIP5 data were used: (e.g. . "The preindustrial state of the land carbon cycle is calibrated against TRENDYv7 and its transient response to CO2 and climate is calibrated against CMIP5 models"). The key thing to communicate in this paper is exactly what data and models were used to calibrate OSCAR, what parameters are being calibrated and what is the sensivity of the model to each piece of information.
We thank the Referee for this interesting comment. As mentioned earlier, all of these aspects have already been published in the reference paper of OSCAR that was cited throughout all the manuscript. In Gasser et al (2017), 49 pages bring the information that you are asking for. As explained earlier, OSCAR v2.2 is described in Gasser et al (2017) while the changes from v2.2 to v3.1 are explained in Gasser et al (2020). In this evaluation paper, no additional or different calibration of the parameters are performed: the model is used in its fixed version 3.1.

However, following your comment, we decided to give more visibility to these explanations in the first paragraph. Besides, we have gone through the text to improve the language.

3 - The exclusion of unstable parameter combinations is understandable - but the conditions for instability need to be more objectively quantified. Plots illustrating the instabilities, and conditions for exclusion would be appreciated, together with a process assessment of why they occur and whether the exclusion process might be biasing the observationally constrained distributions.
We thank the Referee for this interesting comment. We acknowledge that the exclusion process may bias our conclusions. We are now giving more visibility to this aspect in the manuscript.

We had explained that these exclusions were due to the instability of the ocean carbon cycle, hence the described method. Further analysis of this aspect could be performed along the improvement of this module that we mentioned. According to your comment, we are giving more details about the exclusion, in the limit of what is feasible. We explain in more details how we proceed, what it means for the model.

4 - The joint constraint on cumulative carbon uptake and warming results in a tightening of a prior distribution, which was already narrower than the observational uncertainty (Figure 1 (b?) - sublabels are required!). This tightening occurs due to the effect of the constraint on warming, but means that the model ensemble is not plausible sampling solutions with low carbon uptake. This is a potential bias in the assessment of the model distributions, unless we have perfect confidence in the model structure - which we don't. The constraining approach would benefit from having a parameter which allowed for model imperfection in calibration (see McNeall 2016 for an example of this calibration problem and Williamson 2019 as an example of a statistical framework to address it).
We agree with the Referee that the constraining approach could be improved, and the papers that you cite would be of a great help in this sense. We thank the reviewer for providing an interesting lead for further development of the model. We now acknowledge this limit in the paper and cite these papers as a way to overcome it.

5 - It would be useful, thoughout, for plots to show CMIP5 and CMIP6 distributions where available in addition to the OSCAR distributions.

We thank the Referee for this comment. We are adding comparison to the existing literature wherever possible, be it from CMIP5 or CMIP6, or directly from the 6th Assessment Report of the IPCC. It concerns the sections treating the diagnosis of OSCAR. Yet, the section treating the behavior of OSCAR is more for illustration purposes.

We have however elected not to overburden the figures with such comparison, and to limit this type of information to the text or to tables.