Response to reviewers [Revision #2]

"A modeling framework to understand historical and projected ocean climate change in large coupled ensembles"

Yona Silvy, Clément Rousset, Eric Guilyardi, Jean-Baptiste Sallée, Juliette Mignot, Christian Ethé and Gurvan Madec

In this document, reviewer comments have been copied in black, author responses are in blue and citations from the revised manuscript are in green.

We thank the two reviewers for their time and for their final suggestions.

Report #1

Does the historical forced IPSL model contain all forcings (e.g., greenhouse gases, aerosols, and volcanos)? If it does, it should be mentioned somewhere that aerosols could play a role in results over the historical period. Perhaps it could be included in the introductory paragraph on the differences between FAFMIP and this set up, since I don't think FAFMIP include aerosol forcing.

The IPSL historical simulations indeed include all forcings mentioned here, as per the CMIP6 protocol. We have made that clearer in the introduction as suggested.

"Furthermore, the FAFMIP protocol does not focus on the historical period but on an idealized CO_2 forcing, and does not include non-CO_2 anthropogenic forcing agents such as other greenhouse gases and aerosols, which can play an important role on historical climate change patterns (Wang et al. 2016)."

Report #2

The authors have carefully addressed my concern in this revision. I'm delighted to find that more clarifications and results about validation are added, which improve the manuscript. I fully support its publication in EGUsphere after some additional edits.

We are glad that the changes made in the first round of revisions met the reviewer's expectations, and thank them for these additional suggestions.

(The following line numbers are based on the PDF file with tracked changes.)

Comments:

1. Line 111: How about adding “fully-coupled” before AOGCM?

Thank you for the suggestion, we added it.

2. Lines 123-127: Does it mean this approach only works for the model with large-ensemble fully-coupled simulations? How good or bad it would be if only 3~5 (or even 1) fully-coupled realizations are available to use this modeling framework?

As we explained in our first response to reviewers, using a large ensemble is the most appropriate way to extract the externally-forced signal of a model. Averaging over less than 10 members would certainly degrade the estimate of this forced signal. The question of the number of members necessary to extract that externally-forced signal depends on the variable (mean or variance for example), it is a complex question and it is thus not the focus of this study. We show that with 30 members of this model, we can sample the interannual internal variability of the model, and thus obtain an estimate of the forced signal that averages through these variations. There are other ways to estimate the externally-forced signal from a single member, such as fitting a 4-th order polynomial, but this has been shown to be less precise than using the average over a large ensemble (Lehner et al. 2020).

3. Line 134: Are you referring to “ALL experiments” (those ocean-only runs) or just “all experiments”?

We are referring to all the ocean-only experiments. We have clarified this in the text.

4. Lines 145-146: Isn’t climate drift the major factor here?

We do not think so, as climate drift is the same in CTL and in ALL, thus the difference ALL-CTL removes the drift.

5. Lines 160-162: This last sentence seems to be out of place.

This sentence is a follow-up to the previous one mentioning the extraction of the externally-forced signal. We reorganized the sentence to make it more comprehensible.

"Nevertheless, as described above, the experimental design could be applied to any coupled model and its ocean-only configuration, as long as the externally-forced historical and future response can be extracted. To extract this forced response, the historical+scenario large ensemble approach seems to be the most accurate way compared to e.g. fitting a 4th order polynomial to a single member (Lehner et al. (2020)."

6. Caption of Fig. 3: I believe the fluxes shown here from the piControl and CTL are identical. How about rephrasing it to "… from the piControl and used in CTL" to clarify?

Yes, thank you for the suggestion.

7. Caption of Fig. 10: It would be helpful to point out how is the inter-member spread defined.

We have added: "full intermember spread (minimum to maximum)"

8. Caption of Fig. 11/12 and Lines 491-510:
Fig. 11/12 and the associated statements are based on the difference between 2040-2059 and 1850-1899, which reflects the projected climate change mentioned in the title. I'm still wondering what the historical responses look like, such as 1980-2000 minus 1850-1899. Based on Fig. 10, the external forcing seems relatively weak during the historical period. Does the ALL runs behave well during this period?

We refer to our response to comment #4 of Reviewer #2 in the previous response to reviewers (pages 23-26). The ALL experiment in fact behaves better during the historical period compared to the 21st century since it is mostly the forced signal, applied at monthly frequency, which creates non-linear responses, and so the stronger the signal gets the larger the discrepancies become, over time, between ALL and the ensemble mean in strongly non-linear regions.

9. Lines 535-538 and Lines 742-749: I don't think the Southern Ocean is only affected by a little in the absence of the interactive sea-ice model. As I mentioned in #8, the Southern Ocean situation may be worse during the historical time. In the last sentence (Line 537), how about also pointing out the Southern Ocean besides "the Arctic Ocean"? Similarly, the discussion in the last paragraph (Lines 742-749) should also include the Southern Ocean.

As illustrated in Figure 7, 11 and 12, the differences between the coupled configuration and the forced configuration are much larger in the Arctic than everywhere else, including the Southern Ocean. The Arctic stands out, which is not the case of the Southern Ocean (in which the differences are the same order of magnitude than in the global ocean), which is why we mention that the Southern Ocean, compared to the Arctic, seems to be only little affected by the absence of a sea-ice model. As mentioned in the previous comment, the response is in fact better during the historical period than during the scenario period, due to weaker non-linearities.

10. Caption of Fig. 13: Is it the global mean SSS/SST or the average within 60˚S-60˚N?

It is the global average south of 60ºN, we have clarified this in the caption, thank you for pointing this out.

11. References: There are two Silvy 2022 in the references. One is unpublished. If possible, please use a,b to show which one you are referring to.

Silvy (2022) refers to a thesis, deposited online, and Silvy et al. (2022) refers to a published paper in Journal of Climate. We do not have a hand on the format since we used the Copernicus latex template.