EGUsphere-2024-95

Experimental design for the marine ice sheet and ocean model intercomparison project - phase 2 (MISOMIP2)

Reply to reviewers

We would like to thank all reviewers for their detailed and constructive feedback. We have collated our replies to their comments; all original reviewer comments are in black and replies in ***bold italic blue.***

REVIEWER #1

De Rydt et al provide a framework for a modelling intercomparison exercise on ice sheet-ocean interactions in Antarctica. The intercomparison focuses on two key regions of Antarctica: the Amundsen Sea and the Weddell Sea. These two regions are representative of warm and cold ocean conditions and understanding their future evolution is key to predict sea level rise and changes in ocean circulation. The use of realistic configurations as well as the implementation of ice-sheet-ocean coupled models are important steps forward compared to previous intercomparison projects. Very important is also the development of "MIPkit" that allows comparison between model output and oceanographic/glaciological observations (from satellites and in situ measurements). I fully support this modelling exercise, especially because it will allow to look at features/processes in the models that can help explain observed changes and to test the sensitivity of different models to prescribed perturbations based on what we could expect over the next decades.

I have one main comment on the modelling setup and a couple of minor comments.

Main comment

I understand that requiring fixed lateral boundaries is a strong request to the different groups at this stage, especially because it is important to include enough group in the intercomparison project and it is also useful to look at both regional and global models. However, I feel that by having very different lateral boundaries, this could affect interpretation of the model response to external forcing and also internal variability/feedback are likely to change. Maybe some further discussion about this can be useful.

We agree with the reviewer that a Come As You Are Approach allows for significant differences in boundary conditions, model architecture and model physics, which might obscure the origin of model biases and feedbacks. At the same time, it can be challenging to analyze results from model configurations that are asked to fit stringent (forcing) criteria. For the ISOMIP+ and MISOMIP1 experiments, for example, this led to model sensitivities that were far away from their default behavior. Moreover, it is unclear what boundary conditions should be imposed, given the spread in model domain size and biases in existing global simulations. We will better emphasize these considerations in the manuscript, and stress that MISOMIP2 therefore aims to quantify the inter-

model spread and biases for ocean-ice models in their typical configurations. We will identify common and opposing behavior between model setups, which is more challenging but if successful, of greater benefit to the community. We also encourage individual participants to use the MIPkit data to reduce potential biases and optimize their model setups where possible. While we do not discourage contributions with large biases in present-day ocean conditions, we might have to put a lower weighting on those configurations when analyzing the model sensitivity to anomalies in atmospheric forcing and perturbations in ice-shelf geometry.

Minor Comments

- Figure 1: would it be possible to distinguish the location of CTD from moorings?

We will introduce different markers for both categories

- Define CDW and mCDW in the manuscript.

Thank you, this will be done.

REVIEWER #2

This paper documents the design for a series of model intercomparison experiments that aim to probe the realism of and agreement between the models that are used for predicting ocean-driven ice loss from the Antarctic Ice Sheet. The experiments involve simulations under realistic forcing and geometries, and perturbed (but still realistic, i.e. not idealised) forcings and geometries.

I have a very positive overall impression of both the experimental design and its writeup in this GMD paper. The authors have thought carefully about the purpose of their experiments and have clearly specified the many details required. One particularly nice aspect is that the observational data that will be used to test the models has been provided as part of the experimental file sharing. I have no major reservations about the experiments or the paper.

Thank you for the kind words.

Major comment

Given the goals of this MIP, when validating the Weddell Sea ocean model results it seems a significant missed opportunity not to use the borehole mooring and CTD observations from beneath FRIS that have been collected over the years.

We agree that this is an important dataset, and it should be incorporated in the protocol. We will do this in two ways: 1. Participants will be requested to provide vertical profiles of ocean model data at the borehole locations for FRIS and Thwaites. 2. We will investigate the possibility of including a regridded product of the measurements as part of the MIPkit.

On a related note, participants will be requested to provide simulated melt rates at a number of ApRES locations on the FRIS and Thwaites Ice Shelf.

Specific comments

L16: 'with a marine termination' seemed like a strange phrase to me.

We will change this to "...for standalone ice-sheet models with floating ice-shelves have..."

L33: can you cite an 'in prep' paper? It seems unlikely it can be citeable before the present paper is published? So this sentence needs to stand alone, somehow.

We will change this to "X. Asay-Davis, pers. comm."

L40: ditto

Since this work has not yet been submitted, but the lead author is a co-author on this manuscript, we will remove the reference to Smith et al., in prep.

L49 and 50: I was confused by the word 'coupled' here, as I think this is just discussing ice-shelf—ocean simulations.

We will rephrase this paragraph as follows: "However, there is currently limited confidence in the validity of such ice-shelf-ocean simulations for several reasons: 1) important biases remain in the thermohaline structure and dynamical state of the ocean, in particular on the continental shelf, 2) the parameterizations for ice-shelf basal melt used in ocean models is highly tuned and structurally uncertain, 3) no comprehensive comparison between model data and measurements of ocean properties and basal melt has been carried out.

L52: I don't know if it is worth highlighting that the melting parameterisation used in these ocean models (the 3 equations) is itself both highly tuned and structurally uncertain, so anything arising from these models will inherit those uncertainties.

See reply to previous comment.

L59: I would add citations to 10.5194/egusphere-2023-1587 and 10.5194/tc-2023-77 because I think those studies represent advances in the state of the art that are relevant to the setup of this MIP.

Both papers have now been accepted for publication and will be added here.

L86: 'emission-based' -> 'scenario-based'

Will be changed.

L94: Mouginot et al is about ice speed, not grounding lines

Correct. We will refer to Rignot et al 2014, Scheuchl et al 2016, Milillo et al 2019, Milillo et al 2022.

L98: reference typo

Will be corrected.

L102: It would only be a good indication of model robustness if the models were not re-tuned or structurally different between the two locations, but using a common configuration in both regions is not required in the protocol, I think.

A good point, which we will add to the manuscript: "We would consider a good representation of these very distinct environments in a single configuration (or two analogous configurations of the same model) to be a good indication of model robustness."

Section 2.1: I found the naming of the experiments a bit confusing. I feel using numbers is sub-optimal because it gives no clue toward the aim of the experiment. I would also find it easier to have a short name for each experiment. Would the authors consider dropping the numbers and renaming the experiments? To illustrate the point with an example, it could be Ocean-*h, *w, *c, *p, *f or something? (hindcast, warm, control, present geometry, future geometry).

We will change the experiment names to better reflect the aims (* is either A for Amundsen Sea or W for Weddell Sea):

Ocean-*1p -> Ocean*-hind

Ocean-*1vw -> Ocean*-warm

Ocean-*1p-ext -> Ocean*-ctrl

Ocean-*2p -> Ocean*-Pgeom

Ocean-2*f -> Ocean*-Fgeom

L135, and below, e.g. section 2.5: It would save a little work if the participants used the common geometry in their *1p runs, meaning they could skip *2p and having to spin up the model twice. Is it worth highlighting this clearly and recommending participants do this unless they have a good reason not to? Also, if participants do this, could the protocol recommend how they go about it, i.e., do they upload the results twice, or skip one of *1p or *2p, and if so which one, and how do they indicate this is what they are doing?

We agree that some groups can save time by using the *2p geometry for their *1p simulation. We will clarify this in the manuscript. We are however reluctant to enforce the common *2p geometry for all *1p experiments, for two reasons: 1) it goes against the Come As You Are Approach, and 2) groups with an expensive setup or a setup that requires extensive tuning for any given geometry, might not

be able to re-run existing experiments for the different geometry. This would prevent them from contributing to the MIP. Where participants have identical *1p and *2p experiments, they will not be asked to upload the same outputs twice, but they can simply indicate this in the "experiment" variable of the NETCDF global attributes. We will clarify this in the manuscript.

L140: I would imagine coupled models will be likely to be worse than ocean-only models over the historical period, so I would reword this sentence to refer to a 'change in bias'.

We will change the sentence.

L158: Cosgrove does not seem essential to me

We agree. Cosgrove will be removed here.

L159: I feel this needs to be more specific. Is the requirement that the models include the shelf break so that they include all of the shelf? In that case 71S is fine. Or is the requirement that the models include shelf break processes, i.e. undercurrent, in which case I would imagine some decent area of ocean north of the shelf break is needed – 65S? Also, why is there this limitation in the first place? Why can't modelling groups just model whatever parts of the Amundsen Sea they want to? Same comments apply to Weddell Sea domain and shelf break also.

We welcome contributions for any choice of model domain and will remove the reference to 71S.

L168: Why stop in 2022? 1979-present?

We will change this to 1979-present

L179: What observations of sea ice fluxes are available? My personal opinion is that this is a sufficiently different approach that it should be precluded. Or if this is permissible, the choice of surface forcing should be completely CAYA, e.g. including manually created forcings, climatologies, 'normal year' forcing, climate model output, paleoclimate reconstructions, etc etc. I would personally require that the use of a reanalysis and dynamic sea ice model is essential.

We request that participants use interannual reanalysis products to force the atmosphere-ocean boundary and will better emphasize this in the manuscript. This is essential for comparison to observations, and climatological, normal-year and CMIP forcing are not permitted. In terms of sea ice representation, we agree that a dynamical-thermodynamical sea ice model is probably essential, but if a group decided to represent the interannual sea ice variability with a simpler model, we see no reason for excluding this model and hope that its possibly poor fidelity would come as a result of the intercomparison.

L218: Generating absolute forcings from disparate data can be risky in terms of creating unphysical fields, such as negative shortwave, shortwave in winter, relative humidities <0 or >100%, etc. Do users need to be advised on such considerations?

We will add a word of caution about potentially unphysical values when using anomalies.

L231: Naughten paper is now published.

We will include a citation to Naughten et al. 2023.

L260: Is 'early stages' accurate for geometries representing 200y and 300y of strong ocean warming?

We agree and will change the wording to "The aim is to identify and compare the modelled feedbacks between changes in cavity geometry, ocean circulation and basal melt rates for semi-realistic patterns of ice-shelf thinning and grounding line retreat"

Figure 3 caption: final sentence needs correcting.

Thanks, this needs to be "Here the perturbation experiment starts in 2000, and the anomaly is added to the 2000-2018 atmospheric forcing as well as to repeated cycles of the 1979-2018 forcing."

L305: Is it worth explicitly stating that all geographical restrictions on the ocean domain are now dropped?

Yes, here we welcome submissions from models that do not explicitly resolve the ice shelf cavities, but instead use a melt parameterization (such as PICO) forced by 'far-field' ocean conditions, and impose melt fluxes at the front of the closed cavities. We will further elaborate on this in the manuscript.

Following on from above part 1: So this causes a bit of a disconnect between the Ocean and IceOcean experiments, in that models may contribute to the latter without having completed the former?

Correct. We will clarify this in the manuscript.

Following on from above part 2: In general, are there any restrictions on which simulations the participants must perform, or dependencies between experiments? All I understood from the paper was that if participants are completing the *1vw experiments they must also perform the *1p-ext experiments, but even that was not actually explicitly stated. If the participants can perform whatever experiments they like, should that be explicitly stated?

Apart from the *1vw and *1p-ext experiments, there are no dependencies between experiments, and participants are free to choose what experiments they would like to contribute to. This will be stated more explicitly in the manuscript.

General: it would be good to state explicitly whether participants are permitted to upload multiple experiments, and if there is any limit to this (there must be!) or guiding philosophy about what variations might be of interest.

We do not impose restrictions on how many experiments individual participants can submit but will make clear that in the analysis of the full MIP ensemble, we might have to apply a weighted approach to avoid a situation where individual models dominate the mean.

L369: I didn't understand why the ice elevation change and velocities are provided. 1) Is this for model evaluation or initialisation, or what is it for? 2) This seems to go against the CAYA approach, in that participants may prefer different products from those provided. 3) Why host separate online versions of these products when they are already online elsewhere with proper DOIs, version controlled and documented by the data originators?

The ice sheet data were included in the MIPkit for validation purposes only. A weighted average of various velocity and ice-sheet thinning datasets was provided, and regridded to the standard MISOMIP2 grid. Given the CAYA approach, we do not envisage participants to use this data to initialize their models, and will clarify this in the text.

L372: James Clark Ross should be spelt out fully, analogous to Jan De Rydt.

This will be done.

L389: Mark the sections on Figure 1?

We will add output sections to Figure 1.

L402: Important sea ice?

"Important" will be removed from the sentence

L406: Any moorings in the Dotson Trough?

We agree that this is an important dataset, and it should be incorporated in the protocol. We will do this in two ways: 1. Participants will be requested to provide vertical profiles of ocean model data at mooring locations. 2. We will investigate the possibility of including gridded monthly mean product of the measurements as part of the MIPkit.

L502: 'original model' -> 'participating model'

We will change the wording as suggested.

L514: What does 'we prefer seeing dynamical features' mean?

We will clarify this point. Interpolating a coarse resolution model output onto a fine grid imprints the coarse grid meshes (e.g. big rectangles) on the fine grid. When averaging multiple

models, this may hide dynamical structures such as gyres and horizontal gradients, which is why we prefer linear interpolations to the fine grid.

REVIEWER #3

General comments

I thank the authors for their hard work in creating this protocol.

This manuscript describes the protocol for MISOMIP2. This is a follow-up to previous model intercomparison efforts focusing on ice sheet and ocean processes in Antarctica. The goal of MISOMIP2 is to investigate the performance of ocean-only and coupled ice-sheet-ocean models by comparing realistic simulations against observations. The Amundsen and Weddell Seas have been selected to represent warm and cold ocean conditions, respectively. The protocol welcomes a "come as you are" approach with no constraints regarding the model domain, grid resolution, physical parameters, and forcing. This exercise is crucial for testing ice-sheet-ocean models and improving our ability to forecast sea level rise and understand ice-sheet-climate system interactions.

The manuscript is well-written and delivers the necessary information for ocean and ice-sheet modelers to conduct the proposed experiments. My only major comment is regarding possible complications due to the lack of constraints in the forcing of the experiments. For example, in the JRA55-do dataset (Tsujino et al., 2018 - DOI: 10.1016/j.ocemod.2018.07.002), both liquid (river flux of water) and frozen (land ice calving flux) runoff are prescribed around Antarctica and vary with time. In contrast, the CORE2 forcing (Large and Yeager, 2009 - DOI: 10.1007/s00382-008-0441-3) only gives an annual mean river runoff climatology but lacks values around Antarctica. Similar discrepancies may exist in other datasets (MARv3.9.3, ERAint, etc). Differences in runoff forcing can lead to significant changes in, for example, sea-ice concentration (e.g., Tsujino et al., 2020 - DOI: 10.5194/gmd-13-3643-2020), which in turn may affect the ocean and ice-sheet responses. A similar argument can be used for regional models using different datasets as open boundary conditions. Why not, at the very least, restrict the forcing to be employed in these experiments?

The MISOMIP2 protocol is based on a Come As You Are approach, which allows participants to freely choose their preferred approach to treating freshwater fluxes (iceberg melt, runoff from the ice sheet at the grounding line and surface). We agree that different ways of implementing a spatiotemporal distribution and/or liquid/solid partitioning of those fluxes is an important degree of freedom. However, we feel that there is currently no consensus within the community about a preferred approach or dataset, and models currently have very different approaches to adding freshwater fluxes. The ongoing intercomparison project SOFIA (gmd.copernicus.org/articles/16/7289/2023/) starts to address some of these issues. We will clarify how users should treat freshwater fluxes in the

perturbation experiments, as detailed below. The sea-ice model parameters may also have strong effects on surface freshwater fluxes and are usually tuned for a given atmospheric reanalysis, so it is not straightforward to impose a unique reanalysis for all models without intense retuning. Finally, we aim to learn about these choices, e.g. are all models forced by a given reanalysis closer to observations than those forced by another one.

Specific comments

Line 33: "Asay-Davis et al., in prep." - it would have been preferable if the outcomes of ISOMIP+ and MISOMIP1 were published before the MISOMIP2 protocol was released. The lack of documentation of the previous MIP's results may discourage involvement in the current MIP.

We agree, and there is a plan to publish those results.

Lines 214 - 216: how about perturbations in runoff?

We will make it clear that we do not impose any runoff perturbation. First of all, Mathiot and Jourdain (2023) used a Lagrangian iceberg model and did not impose any perturbation of their calving flux, so their total iceberg melt flux is 1100 Gt/yr for both the present-day and the warm experiment. Groups using a Lagrangian iceberg model can use a similar approach. Due to warmer ocean conditions, the iceberg melt pattern may be closer to Antarctica in the future than presently, so the groups imposing an unperturbed freshwater flux at the surface will miss this effect. We nonetheless believe that this effect is small because (1) ice-shelf melting in the warm experiment is more than 10 times larger than iceberg melting in Mathiot et al. (2023) so that most additional freshwater will come from ice shelves, and (2) sea ice production is close to zero in the warm experiment and the stratification therefore stops having a strong modulation role on deep convection. Perturbations in surface runoff (from ice-sheet surface melting) and in subglacial runoff are also excluded as these are not represented in many models.

Lines 270 - 272: The reference densities for seawater and ice might differ among models. Could this cause significant changes in the geometry of ice-shelf cavities when the ice-shelf draft and bathymetry are imposed?

We provide ice-shelf draft, rather than ice thickness, so differences in reference densities should not affect the cavity geometry.

Page 24, Table A1: it might be helpful to include the water mass flux due to liquid and frozen runoff as ocean variables.

According to Caillet et al. (2022), the dominating effect of freshwater fluxes is the buoyancy gain due to the change in salinity rather than to the associated latent heat. We chose to split the freshwater mass fluxes into atmospheric (wfoat), land-ice (flandice), sea-ice (fsitherm)

and correction/restoring (wfocorr) to understand cross-model differences in surface buoyancy fluxes. Splitting them into solid versus liquid contributions is interesting but we prefer not to increase the output size. It should also be noted that flandice mostly consists of solid water (runoff is usually very weak) and that the snow to rain ratio can often be retrieved from reanalyses if needed.

Editorial/Typos

Line 33: “Asay-Davis et al., in prep.” - this citation is missing from the bibliography.

This will be changed to “X. Asay-Davis pers. comm.”

Line 40: “Smith et al., in prep.” - this citation is missing from the bibliography.

Since this work has not yet been submitted, but the lead author is a co-author on this manuscript, we will remove the reference to Smith et al., in prep.

Line 98: "e.g.,[" - please correct the LaTeX code.

Thanks, this will be corrected.

Figure 4: please specify which panels correspond to (a) and (b) or use (left) and (right).

We will add labels to the figure.