We thank the Editor Claudia Timmreck and the three reviewers for their careful evaluation of our manuscript. We found the comments very useful and think that our manuscript will be greatly improved thanks to them. To ensure clarity, the reviewer's comments are written in black and our responses in light blue.

Reviewer #2

The submitted manuscript describes a model investigation of the uncertainties in medium-range projections of the Antarctica Ice Sheet's contribution to sea level. The study focuses on uncertainties related to internal climate variability of climate forcing, both ocean and atmospheric. First, the authors evaluate CMIP6 models, and choose a subset on which to conduct their analysis. They then present historical diagnostics on the available model ensemble for each climate model, extracting the temporal variability of various ocean and atmospheric related variables. Finally, they choose a subset of ensemble members of the SSP2-4.5 scenario for each climate model and use them to force an Elmer/Ice continental Antarctica, resulting in an ensemble of projections for Antarctic Ice Sheet sea-level contribution through year 2100. Results suggest that internal climate variability can affect sea-level contribution, ranging in magnitude from 45-93%, but most of that uncertainty is dominated by atmospheric forcing over ocean forcing. The authors conclude that internal climate variability varies among the climate models, especially for the ocean forcing; therefore, they suggest a strategy for choosing ensemble members that most realistically represent the dominant climate modes of the Antarctic region. They also make recommendations for how to best consider internal climate variability in ice sheet model projections. In general, the methods are well-described and the figures are adequately presented. The analyses and science results are of high quality, and the discussion and conclusion bring up intriguing and relevant points for the ice sheet modeling community.

Overall, I find that this is an interesting study, with important results comparing the effect of internal climate variability due to the ocean and the atmosphere on ice sheet model projections. While results presenting ice sheet modeling projections by themselves could constitute their own manuscript, the authors present much more analysis, including a list of metrics for choosing appropriate model ensemble members to capture internal climate variability. While interesting, these metrics are not the ones used for choosing members for the ensemble results presented. In addition, the authors do not show outcomes that illustrate/quantify the consequences resulting from an ice sheet model using all the suggested updates to their projection procedures. As a result, I find that the addition of these extra results leads to a manuscript that lacks focus. For instance, I think it would benefit the manuscript if some of the secondary analyses were moved to a supplement. In this way, the main manuscript could be dedicated to presenting results specifically on the quantification of the uncertainty in Antarctica's sea level contribution due to internal climate variability. If the authors feel as if the new metrics should be highlighted instead, then a new organization and general story built around those results would benefit the manuscript.

Due to the extensive modification needed, I suggest that major revisions be required before the manuscript is accepted. If the authors work on better organization of the results and on

improving the clarity of their language (per suggestions outlined below), I am confident that this manuscript could result in a valuable scientific contribution to the community. Please see my comments/suggestions/questions below with regards to my major and minor concerns with the current version of the paper.

General comments/questions:

Results show that the ocean internal variability has a minimal effect on the projections, as compared to atmospheric variability or choice of model. If this is the case, why do the designed ensemble metrics mostly focus on evaluating the ocean forcing of ensemble members? More specifically, why is it pertinent to choose members that capture ocean internal variability well if this variability is less important?

We agree and we will better balance and explain the respective roles of the oceanic and atmospheric contributions.

1. The model ranking for Antarctic atmospheric metrics proposed by Agosta et al., (2024) is now referenced (https://doi.org/10.5281/zenodo.11595213). We will therefore revise the paragraph on the choice of the CMIP6 model (§2.1) to give an equivalent weight to the atmosphere and the ocean (we agree that the paragraph currently focus more on oceanic properties).

2. Overall, we agree that internal climate variability mostly affect sea level projections through the surface mass balance (atmosphere), as evidenced by the shaded ranges in Fig. 6. At the basin scales, things can be different. For example, the oceanic internal variability has a stronger effect than the atmospheric variability in some sectors (e.g., basins 1, 5, 9, 10 for the IPSL-CM6A-LR in Fig. 7), and the effects are of comparable magnitude in West Antarctica for the IPSL-CM6A-LR model (Fig. 6h-i).

Do the authors suggest that the climate model ocean representation of internal climate variability lacks skill to the point that using an entire ensemble of forcing does not offer a realistic projection spread? It would improve the manuscript if these questions were considered in the text/discussion/overall story of your paper. It would be even more beneficial to the paper if the authors could support the answers with analysis or results, expanding upon the plots that are already included in the paper.

Assessing the amplitude of the internal climate variability in the ocean is complex. The oceanic internal variability varies greatly depending on the climate models (§3.1 and Fig 3) and is probably underestimated (§4.1), but it is difficult to show this clearly because of the lack of observational data in the ocean over long period. We plan to generalise the assessment of internal oceanic variability carried out in the §3.1 to 15 CMIP6 models, i.e., calculation of across member standard deviation of the 1995-2014 mean potential temperature over the whole continental shelf for the 200-700m depth, in order to better compare the variability of the selected models with other CMIP6 models. The new figure will be discussed in §4.1 and added in supplementary material. The same evaluation will also be carried out on the SMB. For few

regions like the Amundsen Sea, we also have multi-year observations that show a significant variability so that we can consider that a model producing very low variability like MPI-ESM1.2-HR is unrealistic. We nonetheless consider that the variability of IPSL-CM6A-LR may be realistic.

As discussed in your manuscript, climate model ensembles are typically used to represent the spread of model internal climate variability. Forcing the ice sheet model with a large subset of members allows for the propagation of uncertainty due to this variability into projections of sea-level contribution. Here, it is suggested that this might not be appropriate, and that filtering for members that exhibit more realistic variability ("in phase with observed") could be an adopted strategy. Do the authors anticipate that selecting for members would introduce bias into the interpretation of projection uncertainty due to internal climate variability? Is it possible to make runs, or use the runs already completed, to answer this question? (See further questions/comments on this below.)

We will make our recommendation clearer. First, ice sheet models need to be initialised and calibrated to match historical observations. Achieving this would in theory be easier with a forcing from the most realistic CMIP ensemble member, which is why we attempted a selection of the best member. Another reason is that ice sheet simulations can be computationally expensive, and running simulations forced by all members of several CMIP models may not be feasible. However, we agree with the reviewer that once the ice-sheet model is calibrated, the only way to properly assess the uncertainty related to internal climate variability is to force the ice sheet simulations with multiple members. Our study also gives typical relative errors that may be used as relative uncertainty in studies that can't afford to run full ensemble members.

In the conclusion paragraph, there are four listed recommendations for ice sheet model projections. While these are all pertinent discussion points, it is not clear to me why they are included as conclusions of the presented study. More specifically, these four statements - though they may be valid suggestions - are not directly justified by the results shown. It may be that the authors believe that they are, and in this case, please rework this section, so that is clear to the reader how the results map to each of these statements; or perhaps additional figures that better illustrate the connection can be included in the manuscript revision.

We will remove the recommendation on coupled ice-sheet/climate models as it is not clearly demonstrated in our study. Instead, the discussion on ice-sheet/climate coupling will be addressed in section 4.1 which deals with the issue of the robustness of internal variability in climate models. The recommendation regarding initialisation will also be removed, as the topic of initialisation is not directly addressed in the paper.

The other two recommendations stem from the results of our study and will therefore remain in the conclusion but the link with the present study will be clarified.

Specific comments/questions:

Lines 12-14: Please rephrase this sentence. It is awkward and unclear. Also please specify the type of convection you refer to.

We will change *« Conversely, the amplitude of oceanic internal climate variability around Antarctica strongly depends on the climate model as underestimated convection, due to either biases in the sea-ice behaviour or in the ocean stratification, leads to weak mid-depth ocean variability »* to *« Conversely, the amplitude of oceanic internal climate variability around Antarctica strongly depends on the climate model which underestimates convective mixing in the ocean. The latter is due to either biases in the sea-ice production and associated salt rejection or in the ocean stratification that modulates the depth of convective mixing. Such biases lead to weak mid-depth ocean variability »*

Line 15: Please rephrase to something like: "We recommend based on our results that ice sheet model projections consider …" or something similar.

Thanks for the suggestion. We will replace *« We then issue recommendations for future ice-sheet projections: use several members... » with « Based on these results, we recommend that ice sheet model projections consider several members... »*

Line 54: "than" -> "rather than"

Yes, the manuscript will be corrected accordingly.

Lines 58-64: It is difficult to read this list organized in the current configuration. Is there a way to simplify this so it would be easier to digest for a reader, like in a table for instance?

Figure 1 will be moved to Supplementary Material and a table summarising the variables and metrics used for the evaluation will be added.

In the main text, we propose to remove the list and to replace it with a sentence including both atmospheric and oceanic properties evaluation « *The selection of models was first based on the number of members available and on the availability of 6-hourly outputs that were needed to run regional climate projections. It was also based on the model ranking for Antarctic atmospheric metrics proposed by Agosta et al. (2024), and on the model ranking for Southern Ocean metrics provided by the review of three studies which evaluate water masses properties in the Southern Ocean and Antarctic seas (Purich et al., 2021), dynamical properties in the Southern Ocean (Beadling et al., 2020) and bottom properties in the Southern Ocean (Heuzé, 2021). »*

Line 66: Please clarify what is meant by "best" here? Can this be quantified?

For each study, we calculate the RMSE between the CMIP6 model and the observational dataset with which it is compared, for all the variables analysed. The CMIP6 models are then ranked by increasing RMSE.

We will extend the description of ocean properties analysis a bit further than the response to L58-64 by summarising the metric used, the reference dataset to which it is compared and the period over which they are evaluated.

Line 71: "have some kind of", this wording is very informal and difficult to understand. Does it mean that there is a tuning included for the historical? Please articulate this more clearly for the reader.

We will remove the word 'have some kind of' and replace with *« It is also interesting to note that both UKESM1-0-LL and IPSL-CM6A-LR have prescribed vertically distributed ice-shelf melting to ensure the conservation of the ice-sheet mass over the entire simulation, which is known to be important for coastal ocean properties around Antarctica (Mathiot et al., 2017; Donat-Magnin et al., 2021). »*

Line 74: Please expand upon this in your text, including a summary of the period and in what way they compare well to ERA5.

For further explanation on the evaluation of atmospheric properties, the reader may now refer to Agosta et al, (2024) which has just been referenced : https://doi.org/10.5281/zenodo.11595213.

The comparison of atmospheric properties with ERA5 has been expanded to rebalance the explanations of atmospheric and oceanic properties in the selection of the CMIP6 model. We will reshape the existing paragraph as follows *« Agosta et al (2024) evaluate 29 CMIP6 models around Antarctica by comparing their performance with the ERA5 reanalysis over the period 1980-2004 for 9 variables. The models are ranked based on two metrics, which are (i) the mean Root Mean Square Error (RMSE) over the 9 variables normalised by the multi-model RMSE and (ii) the second maximum implausible fraction, which corresponds to the fraction of the surface where the difference between CMIP6 models and ERA5 is greater than three times ERA5 standard deviation. »*

Line 76: Please summarize how these are the best, or add more quantitative language, i.e. the best with reference to what?

See response above for Line 74.

Figure 1: This figure might be better suited for a supplement, since contains more supportive information, based analysis of the climate model runs.

Figure 1 will be moved to Supplementary Material and a table summarising the variables and metrics used for the evaluation will be added.

Line 82: Please include a reference for the friction law.

As this comment appeared several times, we will add the law in the main text for clarity as follows *« The ice dynamics is computed by solving the Shallow Shelf Approximation (SSA) of the Stokes equations (MacAyeal, 1989), assuming an isotropic rheology following Glen's flow law (Glen, 1955) and a linear friction law (i.e.,* $\tau_b = C\, u_b$ *where* τ_b *is the basal shear stress,* C *the friction coefficient and* u_b *the basal ice velocity). »*

Line 85: Please clarify in the text what is meant by curvatures here?

Curvature is the second derivative of the modelled fields (velocity and ice thickness here), i.e., the Hessian matrix. For more explanation, the reader can refer to §2.2 of Gillet-Chaulet et al. (2012):

Anisotropic mesh adaptation is now widely used in numerical simulations especially with finite elements, as it allows to refine the mesh where needed to capture the flow features within a certain accuracy without increasing the computational cost excessively. The method is generally based on an estimation of the interpolation error used to adjust the mesh size so that the discretisation error is equally distributed over the whole domain. It can be shown that an estimate of the interpolation error induced by the meshing is obtained from the Hessian matrix of the modelled field, allowing to define an anisotropic metric tensor at each node (Frey and Alauzet,

In the main text, we will add a parenthesis *« The mesh is preferentially refined both close to the grounding line and in areas where observed surface velocities and thickness show high curvatures (i.e., high second derivative of the modelled field, Gillet-Chaulet et al.,2012) ...»*

Line 104: Please explain in the text more about how this is done. Are there numerical techniques used (inverted?), or is there a procedure designed determine the right correction?

We assume that this comment is about Line 106. We minimise the RMSE between the modelled and the observed ice-sheet mass change for West Antarctica by applying reduction of the friction coefficient, to limit the model drift. A proper inversion is done to obtain the initial basal friction coefficients. Then, these coefficients are adjusted by trial and error to limit the model drift.

We will rephrase as: *« In contrast to (Hill et al., 2023), we do not correct the surface mass balance to maintain a steady state, but we apply a 10% uniform reduction of the inverted friction coefficients to reduce the model drift. For this, we minimise the RMSE between the modelled and the observed ice-sheet mass change (The IMBIE Team, 2018) for West Antarctica. Our configuration overestimates the mass loss trend in the West Antarctica by only 6% but still largely overestimates mass gain in East Antarctica and in the Peninsula (Tab.1). »*

Line 107: “correct” -> pleased rephrase this, as this term is not appropriate to describe model results, and if I am reading the rates right, the WAIS trend is still technically outside of the error bounds.

We will remove the word ‘correct’ and articulate the sentence more clearly *« Our configuration overestimates the mass loss trend in the West Antarctica by only 6% but still largely overestimates mass gain in East Antarctica and in the Peninsula (Tab.1). »*

Lines 109-119: Awkward – please rephrase this last sentence of the paragraph.

We will rephrase as *« However, this bias should not impact most of the analyses presented here, as the projections in response to the CMIP6 climate models are analysed relative to each other. »*

Lines 112-113: Please specify that this statement is for an Antarctic Ice Sheet model CMIP simulation.

Yes, the manuscript will be corrected accordingly.

Line 114: This phrasing is confusing for a reader because the sentence before already implies that you can add them together. I think the point is that we can attribute dynamic ice loss to ocean-forced changes, because the SMB driven dynamics is trivial. Please rephrase.

We will rephrase as *« Antarctic future mass change results from combined effects of surface mass balance and dynamical changes. In standalone ice-sheet simulations, variations in surface mass balance can be attributed to atmospheric-forced changes and dynamical mass loss to ocean-forced changes as SMB changes have little impact on the Antarctic sea-level dynamical contribution over a century (Seroussi et al., 2014; Seroussi et al., 2023). Thus, the effect of atmospheric and oceanic variations on Antarctic sea-level contribution can be analysed separately and then summed to reconstruct the combined effect (Bindschadler et al., 2013). »*

Line 120: Please give more specifics on how these ensembles were chosen. Even though "see next section" is included here as a reference, it is unclear where in the next section this information is included. If so, please note the specific section number for clarity.

We will add the specific section number for clarity.

Line 123: If this statement refers to both types of forcing (atmospheric and ocean), please specify that here, as the sentence is currently vague.

Yes, the manuscript will be corrected accordingly.

Lines 129-135: I do not think this detailed explanation is really needed here. A sentence explaining that MAR output was not available would likely suffice for justification.

We think that this is an important explanation for the ice-sheet community and we would prefer to keep this explanation.

Line 173: Awkward sentence, please rephrase.

We rephrase the sentence « *We now examine the Amundsen Sea more closely as the region is particularly important for the ice-sheet mass loss.* » like this: *« We now focus on the Amundsen Sea, as the region is currently experiencing the largest mass loss in Antarctica. »*

Line 184: "instead" -> something like "as well as" or "in addition to"

Yes, the manuscript will be corrected accordingly.

Line 187: Since one of the conclusions of this paper is related to this point, it would be beneficial to add the not shown plots to the manuscript, instead of just describing them (i.e. for 60 years). Perhaps in a supplement, so that the point can better be made that 60 would be an improvement over 20 years.

A figure similar to Figure 3 with a 60-year period instead of 20-year period will be added in the Supplementary Material to support our arguments.

Line 197: Please note here that this therefore infers more precipitation.

We will add 'resulting in more precipitation' in the sentence *« By 2100 and for the SSP2-4.5 medium scenario, runoff is supposed to remain limited (Kittel et al., 2021), so the SMB is projected to increase largely due to the increased water vapour saturation in warmer air, resulting in more precipitation. »*

Line 198: Please rephrase, e.g., "We therefore focus on variability in SMB components such as precipitation and air temperature."

Yes, thanks for the suggestion. As we also analyse SMB, we will rephrase as *« We therefore focus on variability in SMB and its main components such as precipitation and air temperature. »*

Line 202: consistently -> consistent

Yes, the manuscript will be corrected accordingly.

Line 203: Awkward, please rephrase this sentence.

We will rephrase as *« The largest SMB variability is simulated along the coast of the Amundsen and Bellingshausen seas, which results from the high internal climate variability of atmospheric circulation and air temperature in these regions (Fig. 4d-i). »*

Line 216: Please make these statements more specific, e.g., … explained by "grounding line change and dynamic response" of these glaciers.

We will rephrase as *« The West Antarctic positive SLC is mostly explained by the grounding line migration and the dynamical response of Pine Island and Thwaites ice shelves (~3 cm in Fig. 5c, basin 11) as well as Getz ice shelf (~1 cm, basin 10). »*

Lines 226-227: relative -> related? (both instances in the sentence)

Yes, *related* is the right word. The manuscript will be corrected accordingly.

Line 252: although -> even though

Yes, the manuscript will be corrected accordingly.

Lines 265-267: The wording here is confusing, please rephrase.

We will remove the argument of GHG scenario in the sentence and focus mainly on the present-day drift and we will mention the sources of uncertainty described in Seroussi et al., (2023), i.e., uncertainties in the physics of the ice-sheet model, the choice of climate model and uncertainties associated with ice-climate interaction (melt parameterisation and calibration).

Line 269: It would be helpful to also note in the text that for the 20 years in question, the Antarctic Ice Sheet of reality appeared to be in a stable state, as opposed to its state under the random phase of a climate model you mention that would occur during these same "years" in a climate model.

The ice sheet of reality was not necessarily in a stable state, e.g., it lost mass in the Amundsen Sea sector. But the random phase does change the mass evolution. We have not added anything on this point.

Line 270: Please clarify here what is meant here by high variability of 20-year means.

We will replace with « *wide confidence interval on a 20-year mean* ».

Line 279: "applying" -> "to apply"

Yes, the manuscript will be corrected accordingly.

Line 290: The previous paragraphs justify why this would make sense to do, but please add some discussion about how one would go from that previous logic to each of the outlined decisions here. More specifically, based on this list of assessments, as a reader I can gather that one should choose the members that have the best representation of important modes of variability known to affect the ocean and atmosphere in/around Antarctica. Please walk the reader through why these key internal variability metrics are chosen, i.e. what it is in the results that can point to each of these as a justified criterion for ensemble members.

The discussion section on the identification of the best member will be revised. To facilitate reading, Figure 8 and the list of associated metrics will be moved to the appendix and similar analysis will be performed with UKESM-0-LL to get more robust conclusions.

In the revised paragraph, we will specify that the metrics considered were chosen to:

– ensure a good representation of the mean atmospheric and oceanic states. We selected variables directly used to drive the ice sheet model, i.e., the SMB for the atmosphere and temperature for the ocean. We focused on the ocean temperature in the Amundsen sector as the region experiences the current main mass loss and CTD profile data are available for a relatively long period from 1994 to 2018 in this area.

– ensure a good representation of the amplitude of oceanic variability using the same observational data described in the previous paragraph. We did not evaluate the variability of SMB since it has been relatively stable in recent years and there is no observational data.

– ensure the best representation of important modes of variability known to affect the ocean and atmosphere in/around Antarctica. We focus our analyses on the SAM and TPI index.

– ensure a phasing of internal variability with observations, which could be important for future detection/attribution studies and for projected Antarctic sea-level contribution. We chose two variables, sea-ice concentration and the presence of warm periods on the continental shelf of the Amundsen Sea to provide insights on the phasing of internal variability.

It should be noted that this part of our study remains exploratory and the choice of variables and metrics is indeed very subjective, which is part of the caveats that we discuss.

Line 344: Please remind the reader here that this is only for the IPSL members.

Following the response to the previous comment, the research for potential best members will be conducted only for both IPSL-CM6A-LR and UKESM1-0-LL.

Line 354: Would the results (projections and spread) change if one were to follow the newly created rankings and chose a different group of members for the analysis? I realize it might not be possible to do new runs to completely answer this question, but are there subsets of runs that are already completed that can be used (i.e. redo analysis of runs with top half vs. bottom half of members?). Or even still, in your discussion, is it possible to use the results presented and extrapolate the results to suggest how much it would matter to the projections to use the ranked members instead?

We believe that we do not have enough simulations to divide our set into two subsets (only two members in each subset for UKESM and IPSL). However, we can comment on whether the best-ranked members tend to have a strong/weak dynamical or SMB contribution.

Line 367: Please specify what is meant by convection here, for example where, and at what depth?

For clarity, we replace '*convection*' by '*ocean convective mixing on the continental shelf*'. The entire water column is affected.

Line 368: Please rephrase, i.e. changing -> "using the following practices to remove the effects of internal climate variability on projection results…"

Based on the comments of all the reviewers, we decided to only keep two recommendations in the conclusion, so this part has been reshaped.

Lines 369-372: Here, metrics to find ensemble members that are in phase with observed variability are described. Please comment on the consequences of filtering these members. In a way, this could defeat the purpose of using a large ensemble, because it discounts variability that is intrinsic within a climate model (even if it is a variability that is not realistic or wrong). Is doing a filtering justified because results suggest that the ocean models are so wrong (i.e. the various members are all over the place in terms of variability and not at all "realistic"), that they no longer offer trustworthy projections? Is it important to rank climate models themselves based on metrics of variability we can derive from their ensemble, in order to tell the community which ones might have completely unrealistic internal variability? Adding these concepts to the discussion would broaden and enhance the scope and impact of the presented work.

Thanks for the valuable suggestion. First of all, a justification for filtering members in the calculation of metrics is that ice sheets act as low-pass filters, they are not very sensitive to interannual forcing. We do not suggest that the variability is wrong in all ocean components, this conclusion only holds for MPI-ESM1.2-HR. Second, we agree that it is important to show

where these 3 models stand in terms of internal climate variability, and we will add a comparison to a larger set of CMIP6 models in the revised manuscript.

Line 374: Please add some text on why this is so, i.e., add a "because …". It would be helpful to clearly state your reasoning here. For example, 1) climate models show important modes longer than 20 years, and/or 2) the 20 years in an observational period that may seem appropriate in "reality" may not align with an appropriate period within a climate model.

We plan to develop a more robust argumentation for this point in a discussion section.

Line 377: "but not in the observations", following this, please add something like "and therefore introduce a different source of bias."

Yes, the manuscript will be corrected accordingly.

Line 382: "Ice-sheet models should" -> something like "To capture the full uncertainty due to internal climate variability, ice sheet models would ideally be …"

Yes, the manuscript will be corrected accordingly.