Response to the reviewers

We thank all referees for their thoughtful and thorough reviews of our paper. In the following we address their concerns point by point. Throughout this response to review document referee comments are provided in black font, our response comments are provided in red font.
To accommodate the main comments of the reviewers, we have made the following substantial changes to the manuscript:

- Expanded the model description, with particular emphasis on the energy balance models (EBMs) and their respective formulations.
- Expanded the Discussion section to place our findings in a broader scientific context and to strengthen the links to previous studies.
- Added further discussion of methodological considerations and limitations, including the representation of ice–ocean interactions, the temporal resolution of atmospheric forcing, and the influence of initial ice-sheet geometry.

Reviewer 1 – Jeremy Fyke

Reviewer Point P 1.1 — This study explores the reasons for substantially different long-term projections (i.e. past year 2100) of Greenland Ice Sheet evolution under anthropogenic forcing, using an ensemble of coupled ice sheet/climate models of various complexities. This paper provides a good example of intercomparison value: through a structured experimental design, among other results it highlights an important model process (SMB modelling) for the question at hand (long-term GrIS sea level rise contribution). This in turn should motivate the community to prioritize work on this arguably key process, rather than potentially less influential ones.

While I concur with conclusion that SMB/climate are very important to long-term GrIS trends, my primary general comment is that I'm not sure the present manuscript demonstrates this quite as conclusively as claimed. Particularly, because the sensitivity testing nominally meant to disentangle the climate/SMB/topography factors on GrIS response seem limited to 2100 and based on SMB model-only results, so results inferring their long term impacts on GrIS mass loss (a.k.a. sea level) remain somewhat just that, inferred. For example, over the 1900 remaining years of the multi-millennial simulations shown elsewhere in the paper, it could well be that topography differences emerge as a primary factor via the height/SMB feedback, or, that late-stage ice dynamical effects (e.g. shifting surface gradient patterns) become important. So in general I think greater justification is warranted during revisions, to try and highlight caveats related to using SMB-model-only, to-2100 behaviour to infer SMB/climate/topography impacts on multi-millennial timescales. That said, authors are welcome to push back on this general sentiment if they feel it is off-base.

Reply: We thank the reviewer for the overall positive assessment of our manuscript. We agree that the caveats of our experiments should be highlighted more clearly. To address this point, we have tempered the language throughout the manuscript and provided a more explicit discussion of the advantages and limitations of our model intercomparison approach.

Reviewer Point P 1.2 — Paper theme, and title: I think topic of this paper is, at its core, a targeted exploration of model uncertainty on long term GrIS change, using some sensitivity testing to do this. Maybe saying this clearly, in the title and elsewhere in the Abstract/Intro, would give readers a clearer cognitive expectation of the study contents. So for the title, instead of 'Unravelling the source for diverging...' what about 'Partitioning model uncertainty sources behind diverging...'. Or even more directly: 'Surface mass balance model formulation as a primary source for diverging...'?

Reply: Thanks for the suggestion. We have changed the title to: "Surface mass balance as a primary source for diverging long-term ice loss trajectories within the Greenland Ice sheet Coupled Model Intercomparison Project (GrICMIP)".

Reviewer Point P 1.3 — L2: 'Since 1990...' – suggest referring to an earlier, climatological period, outside of the transient historical period.

Reply: We rephrased this to read: "Throughout the historical period from 1850–1990, the Greenland ice sheet (GrIS) remained close to equilibrium, but since 1990, its overall mass loss has increased from a near-zero contribution to ∼250 Gt/yr, or ∼0.7 mm/yr sea-level equivalent (SLE), in the 2010s [Mankoff et al., 2021]."

Reviewer Point P 1.4 — L31-L38: 'Almost exclusively...', '...in its infancy...', '...have been limited...' I think the authors could possibly detune these claims a little bit. There has been a fair amount of model development/science over the years that falls into the categories described in these sentences. I don't think these somewhat subjective claims on the history of the research field are necessary to frame the value of the current paper.

Reply: We agree and have tempered the language accordingly.

Reviewer Point P 1.5 — L40: 'By a factor of four...' – would it be possible to convert this into a 0-order estimate of the difference in worldwide coastal inundation and/or displaced population (based on current population distribution)? This restatement would make this 4x factor immediately more relevant to decision making audiences, if formulated reasonably.

Reply: It is difficult to estimate a global sea-level number from two numbers at the end of a 300 year simulation. To nevertheless emphasise the magnitude and relevance of the difference, we have added the following statement: "This difference corresponds to a GrIS mass-loss rate of ∼2853 Gt/yr, equivalent to approximately 8 mm/yr SLE, or more than an order of magnitude larger than the ∼0.7 mm/yr contribution from the GrIS observed in the 2010s [Mankoff et al., 2021]."

Reviewer Point P 1.6 — Section 2: I think at the start of Methods, a paragraph or two is needed to very clearly restate, at a high level and in relatively plain language, the overall goal of the experiment, a statement of the experimental design, and the rational why the current experimental design achieves experimental goal. This is already touched on in the Introduction, but I think readers would benefit from having it again very directly and clearly restated at the beginning of Methods, so they have this context solidly in their mental RAM before they enter the 'weeds' of the different models and specific experimental protocols.

Reply: We have added a paragraph at the beginning of the Methods section to guide the reader through the experimental roadmap of the paper. It reads: "The study is structured around two complementary sets of experiments. First, we perform coupled future-projection simulations to characterize differences in the simulated climate and ice-sheet evolution across the participating models and identify the key processes behind these differences. Second, we use stand-alone EBM simulations to assess the contribution of individual processes to the differences identified in the coupled simulations. Thus, the coupled simulations establish where and how the models differ, while the EBM experiments help us assess why they differ. Because the EBMs and their interactions with the climate and ice sheets are central to this analysis, we provide a more detailed description of these model components below."

Reviewer Point P 1.7 — Section 2.1.4: SMB is presumably quite dependent on the representation of snow/firn model in each of the models (e.g. multi-layer mixed-phase? Hydrologic/liquid water representation? Radiation-penetrating? Firn resolving?). More details on this, in this section, would seem a valuable addition. More generally, would it be possible for the authors to disentangle the effect of the EBM modelling differences and the snow/firn modelling differences, on the final result?

Reply: We agree with the reviewer and have expanded the description of the EBMs to better reflect their central role in the study. We have also synchronised the descriptions to facilitate a more direct comparison between the models.

Reviewer Point P 1.8 — Section 2.1.4: If different SMB models A and B were 'perfect' representations of real SMB processes (C), they would all also theoretically be functionally identical to each other (if A=C and B=C, A=B). So, following this line of reasoning, differences between SMB models stem from the different, necessary choices that each modelling group took to simplify infinitely complex reality down a tractable computer-based SMB model code. For this reason I wonder if this section could explicitly highlight, to the author's views, the primary simplifications/caveats of each SMB model. Definitely not to criticize any of them per se. But rather just to help objectively provide the reader with the potentially primary reasons behind the different SMB model (and resulting long term sea level) results. And perhaps point to community-wide priority improvement areas.

Reply: We agree with the reviewer and have expanded the descriptions of the EBMs to clarify their underlying assumptions, key simplifications, and limitations. We have also synchronised the descriptions to facilitate a more direct comparison between the models and help readers understand how differences in model formulation may contribute to differences in SMB estimates and the resulting long-term sea-level projections.

Reviewer Point P 1.9 — Section 2.1.4: AWI-ESM dEBM description: 'the melt period is defined as the faction of the diurnal cycle during which incoming shortwave radiation exceeds outgoing longwave radiation': but I am confused since I can easily imagine cases where no melting occurs, even if the former exceeds the latter (specifically, at very cold baseline temperatures, well below 0C). Should 'melt period' be called 'potential melt period' instead perhaps? Additionally, it's not clear how PDDs are used, in this EBM architecture. This all presumably points to confusion on my (i.e. the readers') part instead of dEBM modelling issues. In which case, perhaps a clearer description of the dEBM approach is needed, instead of just references to a secondary paper.

Reply: The reviewer is correct that the melt period, as defined, does not by itself imply melting. dEBM evaluates the surface energy balance under the assumption that the surface is at the melting point. The melt period is the part of a clear-sky day during which incoming shortwave radiation exceeds the outgoing longwave radiation of such a surface. It is derived analytically from the local, time-dependent solar geometry and the diagnosed atmospheric emissivity. Melting during this period occurs only if the energy balance of the melt period is positive. In addition, melt is ruled out entirely where the monthly mean near-surface air temperature is below -6.5°C. Very cold conditions such as those the reviewer describes therefore yield no melt. We agree that "potential melt period" is the more accurate term and have adopted it in the revised text. PDDs are not used as an empirical melt factor in dEBM. They only serve to estimate the mean near-surface air temperature during the potential melt period, which enters the temperature-dependent longwave and turbulent heat flux terms of the melt-period energy balance. Following Calov and Greve (2005), this temperature is derived from the monthly mean near-surface air temperature, assuming normally distributed sub-monthly temperatures with a constant standard deviation of 3.5°C. This accounts for the fact that daytime temperatures during the melt period can exceed the monthly mean. We have rephrased the corresponding sentence to make this explicit.

Reviewer Point P 1.10 — L179: '. . . are performed with the different EBMs in standalone mode.' So, to be clear, the core of the study is using static EBMs, uncoupled to either climate or ice sheet. In this case think it'd be worthwhile to make this very clear – and potentially, worth reweighting the methodology description towards a much deeper description of the EBMs, in lieu of the coupled climate models more broadly.

Reply: We agree and have significantly expanded the EBM model descriptions, as noted in our previous responses. We have also clarified, at the beginning of the Methods section, the overall purpose of the stand-alone EBM simulations.

Reviewer Point P 1.11 — L181. Somewhat by definition, the energy balance at the terrestrial surface (ice-covered or otherwise) will impact overlying atmospheric conditions in reality (and coupled climate models). This self-consistent nature is, arguably, a big benefit of coupled modelling. Imposing fixed atmospheric conditions developed with one coupled models setup (in other words, where the atmospheric conditions have a 'signature' resulting from existing above the particular EBM formulation) onto another EBM as prescribed forcing breaks or 'decouples' this link. As a conceptual example of an impact of this: the transfer of atmospheric heat content into latent heat absorption to melt ice cools the near-surface air, in reality or a 2-way coupled model. But this atmospheric cooling effect - reflecting conservation of energy - is presumably lost in an uncoupled setup where the air temperature is prescribed. In other words in this case, in this situation, some heat is effectively created, presumably resulting in some level of over overestimated localized melting..? Could the authors potentially describe or at least speculate what impact or signature this 'decoupling' could have on their results?

Reply: In the real world, melt-induced atmospheric cooling can feed back onto surface melt. However, in our three models, the coupling between the EBM, ISM, and climate model occurs annually. In addition, the atmospheric model does not receive information directly from the EBM. Overall, this means because both the coupled simulations and the stand-alone EBM simulations do not represent this process, this should not cause any differences in the simulations.

Reviewer Point P 1.12 — Use of different scenarios: different SSP scenarios add an extra dimension to the overall analysis (including nonlinearities due to model and scenario-specific AMOC collapse), that I wonder might provide more confusion than value? Is including the complication of multiple scenarios actually necessary at all to meet the primary objective of the study? It seems like the approach of just focusing 100% on SSP245 (e.g. as in L266 and onwards) and completely excluding results from the other scenarios entirely, could be expanded study-wide, to achieve essentially all of the study's goals. I.e. 'Here we focus exclusively on results obtained using SSP245, noting that extensive sensitivity testing resulted in qualitatively similar results...' Would this be feasible, in the author's view?

Reply: We believe the additional SSP scenarios provide valuable evidence that, although the magnitude of the simulated changes varies, the overall patterns remain robust across emission scenarios. We therefore consider them an important part of the analysis and would prefer to retain them in the manuscript.

Reviewer Point P 1.13 — General: ensure that all acronyms are well defined. And in more general, I suggest doing an overall edit while taking the naïve perspective of a new reader. The authors may find, in emulating this mindset, that many short-hand terms that are intrinsically familiar to them might benefit from a bit more plain-language description ('spin-up procedures', 'dome structure', for example).

Reply: We have revised the manuscript to ensure that acronyms are defined at first mention and to improve clarity for readers who may be less familiar with the terminology. In particular, we have reviewed the text to reduce unnecessary jargon and added further explanations of technical terms where appropriate.

Reviewer Point P 1.14 — Figure 2: A really interesting graphical depiction of model uncertainty. I think it's worth noting that there are other, non-ice-sheet examples in climate modelling that display similar levels of specific intermodal heterogeneity ('lack of consensus') as part of coupled future projections. Like, changes in overall wind strength patterns, Amazon response, El Nino variability... I almost wonder if making a 'metapoint' in this paper, that the differences in GrIS response across models under the same scenario is in many ways consistent with sometimes quite significant differences in non-ice-sheet metrics evaluated in other multimodel analyses (including some citations to relevant, analagous studies). In other words, it's not GrIS alone that does quite different things in different coupled climate models, for sometimes hard-to-understand reasons...

Reply: We agree that this is an interesting analogy. We added the following to the discussion: "More broadly, intermodel differences in projected GrIS evolution reflect a wider challenge in climate modelling, as substantial differences also occur in projections of other components of the climate system, including ENSO variability and the Amazon carbon sink [Beobide-Arsuaga et al., 2021, Erickson and Patricola, 2023, Mastropierro et al., 2025]. Such differences partly reflect model-dependent representations of the processes and feedbacks governing the climate system. The spread in GrIS projections should therefore be considered in the broader context of uncertainty in coupled climate projections."

Reviewer Point P 1.15 — Figure 3: the increase in calving anomaly here could be confusing to readers..? Essentially, the positive anomaly values here are indicating a less negative flux as calving

reduces due to more land-terminating ice (right?). And this is being compared to less positive SMB budgets over time (the solid lines). I wonder if a clarification in the caption, or, a splitting into two plots (where calving flux is considered a positive number) would be useful.

Reply: We agree that this deserves more elaboration. We amended the Figure caption to read: "Time series of anomalies in ice-mass budget components relative to pre-industrial (PI) conditions for all models: (a) ice-mass budget components, as defined in Eq. 1, and (b) individual surface mass balance components of the Greenland Ice Sheet for the SSP2-4.5 emission scenario. The abbreviations are as follows: smb, surface mass balance; bmb, basal mass balance under ice shelves. In (a), calving is represented as a negative contribution to the ice-mass budget; therefore, a positive calving anomaly indicates a reduction in ice loss due to calving relative to PI conditions. The smb time series in (a) and (b) are identical for each model. Model output is shown as decadal means."

Reviewer Point P 1.16 — Figure 3: it appears that accumulation is periodically net-negative past year 3000 for AWI-ESM and MPI-ESM. This seems like at least an integrated SMB accounting error – i.e. how can there be negative accumulation (that sort of implies upwards snowfall)? It perhaps points to an issue in the determination of different SMB terms for this plot..?

Reply: We thank the reviewer for pointing this out. The plotted values are anomalies relative to pre-industrial (PI) conditions, rather than absolute accumulation rates. Thus, negative values indicate that accumulation has decreased relative to PI conditions, not that the absolute accumulation rate is negative. We have clarified this in the figure caption.

Reviewer Point P 1.17 — L356-L357 – '. . . it becomes evident that different surface topographies cause the smallest differences in SMB.' Do the authors mean, essentially the original spun-up topographies, have little effect on SMB spatial distribution at year 2100? This could be a bit confusing to interpret for readers, particularly as they ponder how increasingly divergent topographies (in response to different SMB fields) contribute in turn to even more divergent SMB mass trends, via the positive (amplifying) height/SMB feedback. Perhaps the authors could add some nuance here, to reflect this time dependency..?

Reply: This is a good point. We have added the critical information that we are referring to the "initial topography".

Reviewer Point P 1.18 — Section 3.5: I suggest separating into three different subsections (or at least, 3 different paragraphs). Especially since this seems like a central aspect of the study. It is really hard to parse this single 40-line paragraph, that covers all three sensitivity assessments.

Reply: We agree and have divided it into three separate paragraphs now.

Reviewer Point P 1.19 — Surface air temperature is actually not, itself, the direct driver of melting. Instead, it is a key regulator the energy fluxes between the ice sheet surface and the atmosphere; and, it is these fluxes that cause melting. Really, the advantage of EBM modelling, is that it captures this temperature/melt relationship in a physically consistent way, rather than relying on empirical relationships like PDDs. To that end, given the advertised importance of using EBM modelling, I think the paper could use a bit more nuance when tying GrIS changes to changes in air temperatures specifically. This nuance might uncover why, exactly, the SMB

models are behaving differently under similar climate/topography conditions, as they 'transform' near surface air temperature into the melt response, via the EBM modelling structure. In general, I feel like this investigation would be a valuable contribution if it were possible, since it would point to priority areas for EBM modelling improvement for better sea level rise projections.

Reply: We agree that this is an important distinction to make. We have revised the Discussion to clarify that near-surface air temperature does not directly determine surface melt.

Reviewer Point P 1.20 — ECS is a measure of global average sensitivity. But the additional, equally important 'bridge factor' that relates global ECS value to high latitude conditions that matter for GrIS, is polar amplification. For example, one can imagine a similar GrIS response, for reasonable high ECS/low PA versus low ECS/high PA models. In fact, this has been demonstrated in a controlled setting (https://link.springer.com/article/10.1007/s00382-014-2050-7, full disclosure, I am lead author of this somewhat dated paper). For this reason (amongst others), I'm not entirely surprised that, for example, models with similar ECS display quite different GrIS responses, and leads me to agree with the statement 'underlining the critical role of the characteristics of the climate model for the future evolution of the GrIS'.

Reply: We thank the reviewer for this insightful comment. We agree that ECS alone does not capture the regional temperature response relevant to the GrIS, and that polar amplification provides an important additional perspective. We have therefore calculated the polar amplification factors for our three models under SSP2-4.5 and incorporated the results into the Discussion. CLIMBER-X exhibits the lowest amplification factor, consistent with its weaker GrIS mass loss, although the comparison across all three models indicates that amplification alone does not explain the differences in ice-sheet response. One important consideration is that polar amplification is generally stronger in winter, e.g. due to sea-ice loss, whereas the surface mass balance of the GrIS is particularly sensitive to summer temperatures, which govern surface melting. Consequently, neither ECS nor the combination of ECS and polar amplification fully captures the ice-sheet response. We now discuss these results alongside the role of ECS and the model-dependent processes governing the regional climate response and surface melt.

Reviewer Point P 1.21 — Climate feedbacks (including those related to ice sheet evolution) 'turn on' as a function of the applied perturbation. So in this sense, scenarios with strong climate perturbations (e.g. SSP5-8.5) should display strong feedback response, and thus be reasonable places to investigate climate-ice sheet interactions (contra the hypothesis on L436, if I understand it correctly). More generally: if the goal is to study these processes, I would argue that simple step-function perturbation forcing scenarios (e.g. instantaneously doubled CO2, or 3x CO2 as in the above-referenced paper) are the most appropriate, because there is less conflation between gradual forcing changes, and ice sheet response. Recognizing that perhaps these exact experiments are not available for the current study, it might be nonetheless worth mentioning as a way to further understand relative roles of topography, SMB formulation, and climate, in long-term response.

Reply: We do not fully agree with this statement. It implies that stronger perturbations, or in this case higher-emission scenarios, necessarily increase the signal-to-noise ratio of climate–ice sheet interactions relative to the direct effects of global warming. We argue that this is not necessarily the case, particularly for interactions that reinforce the direct warming response. For example, under strong-emission scenarios, global warming itself is expected to weaken the AMOC by reducing ocean buoyancy. The additional

freshwater input from ice sheet melt may have a comparatively minor effect on this response. Conversely, the perturbation must be sufficiently strong for the response to emerge from internal climate variability and model noise. As demonstrated by Testorf et al. 2026, isolating new feedback mechanisms may even require suppressing stronger interactions to make the signal of interest detectable. Thus, we argue that the choice of the optimal perturbation magnitude is more nuanced than simply opting for the strongest possible forcing. We have amended the discussion to reflect this more nuanced view.

Reviewer 2

Reviewer Point P 2.1 — Schanwell et al. have conducted simulations of the future evolution of the Greenland ice sheet (GrIS) and its contribution to global mean sea-level rise for the next 2000 years under three different climate change Scenarios using three different climate models that cover a (limited) spectrum of complexity and which include coupled ice sheet components. Their primary analysis is concerned with understanding the significant differences in this evolution as it is simulated by the different models, which they attribute primarily to differences in modelling the climate and its influence on how ice sheet surface mass balance is represented, downplaying the role in such projections of the initial state of the ice sheet model itself. The paper is quite clearly written - although sometimes a little too concise for my preferences - and the figures are generally appropriate to support the stated scope of the analysis - although the multi-panel figures are sometimes too small in hardcopy when printed out on A4 pages. I think the conclusions drawn could be caveated a little more with regard to the timescales over which they have identified these dominant factors and that the role of ice dynamics has not really been isolated independently from the climate and SMB representation in the model. With some more careful statements around the conclusions and some more contextual detail given in the model descriptions and description of results I think this would make a useful contribution to the emerging literature on coupled climate - ice sheet modelling.

Reply: We thank the reviewer for their overall positive assessment of our manuscript. We agree that some aspects of our discussion did not sufficiently acknowledge the advantages and limitations of our experimental design. We have therefore moderated our wording and added appropriate caveats throughout the manuscript to better reflect these considerations.

Reviewer Point P 2.2 — Although the paper does quote numbers for the sea level contributions from each model under each scenario, not a lot of emphasis is placed on these, preferring instead to try to understand the cause for the differences, which I think is a good use of simulations like these. Their main conclusion, and thus the focus of pretty much all that is written in the paper, is that the local climate evolution and SMB representation are the main factors. This point is made clearer by the fact that the experiment has been done in the same way across the three models, and in general the paper is a nice demonstration that model structural issues in representation of climate and downscaling (ie EBM) to allow ISM coupling need as much, if not more, attention than the coupling mechanism and the ice sheet model itself. That said, it would have been nice to see the MIP structure exploited further to understand the difference in retreat patterns seen in each of the models.

Reply: As changes in the ice sheet are primarily driven by changes in the surface mass balance (SMB), differences in the spatial structure of the SMB fields explain the distinct retreat patterns across the three models. We have added the following sentences to the manuscript to clarify this relationship: "Interestingly, the ice-thinning patterns for each individual model are very distinct. These differences in retreat patterns can be attributed to the distinct spatial structures of the SMB fields simulated by the three models, which lead to different spatial distributions of surface mass loss and, consequently, different patterns of ice-sheet thinning and retreat." We also added a paragraph to section 3.5 that discusses the potential influence of the three tested sensitivities on the ice-sheet retreat pattern.

Reviewer Point P 2.3 — I find the paper quite brief in a few aspects where I think more contextual detail would help with a readers' understanding (eg model description). I also felt it could do with more context of the ice sheet simulation in each model including some evaluation of performance over recent historical period to help readers get a more complete view of what happens in the simulations

Reply: We thank the reviewer for this suggestion. We have added a figure to the Appendix comparing the simulated ice thickness in the three models with BedMachine observations, including the corresponding differences. We also discuss these differences in the revised manuscript.

Reviewer Point P 2.4 — It would be good to note more clearly in the design of the factorial experiment, where the usable results come before the ice sheet has significantly evolved in each model, that they have not untangled the role of ice dynamics compared to the simulation of climate - this is an important limitation that I think could be made much clearer. The authors could also be more careful about stating the timescales over which unimportance of initial condition is true, since there is evidence that in the first century or so the ice sheet initial topography does play a significant role.

Reply: We have explicitly clarified that ice dynamics are not the focus of our study and qualified the conditions under which the initial topography is of lesser importance.

Reviewer Point P 2.5 — Title: I'm not sure you can 'unravel' something that doesn't explicitly have multiple parts. "Unravelling the sources of diverging ..." perhaps

Reply: We have changed the title of the paper according to Reviewer 1's suggestion to: "Surface mass balance as a primary source for diverging long-term ice loss trajectories within the Greenland Ice sheet Coupled Model Intercomparison Project (GrICMIP)"

Reviewer Point P 2.6 — line 56: In general there's a lot of this section that requires readers to go dig through the original model description papers to find detail beyond the submodel names and resolutions. In many places I feel like it would be useful for readers if this section (or at least the Appendix) had more detail of factors affecting relevant physics features and intepretation of how those details are likely to affect the model's simulation eg sliding laws, going beyond the simple naming of the sliding law in use in table A1.

Reply: As requested by Reviewer 1, we have expanded the model description detail of the EBMs and also tried to synchronise the model description. In addition, we have also clarified, at the beginning of the Methods section, the overall purpose of the stand-alone EBM simulations.

Reviewer Point P 2.7 — line 111: the MPI-ESM section states that the AIS is prescribed in this study, but also makes specific reference to the resolution of the ice sheet model grid for Antarctica which makes the status of how the AIS is handled in this configurations a bit confusing. Could you clarify?

Reply: We agree. We have removed the confusing resolution statement.

Reviewer Point P 2.8 — line 120: a similar comment to the one on general model descriptions above, but it's even more important here to describe all the relevant detail in how the snow and firn are modelled and the surface mass balance is calculated in each case. The three descriptions could be harmonised to cover the same aspects in the same way, rather than being three seemingly-independent descriptions. Eg, in each section a list of exactly which atmospheric variables are used, how refreezing is handled - which is touched on in the CLIMBER-X description but absent in others. As with the comment on the general model description, some interpretation of the relevant limitations and implications in the following simulations of the parameterisation choices wold help interpretation.

Reply: We agree with the reviewer and have substantially expanded the description of the EBM formulations and harmonised the presentation across the three models.

Reviewer Point P 2.9 — line 167: Again, at least little more detail on the construction of the extension of the SSP scenarios to yr 4000 and to what degree the idealised nature of what has been done may impact the relevance of the results to more realistic futures would be welcome, rather than diverting the reader to the Supplementary Material of a previous publication. The atmospheric CO2 concentrations that result from the CLIMBER-X emissions-driven simulations are shown, but no information is given here about what emissions have been used to create them.

Reply: We have added some more details on the extended emission-driven SSP scenarios: "As forcing, we utilized time-varying CO_2 and CH_4 levels computed in separate emission driven CLIMBER-X simulations for the different scenarios. Future anthropogenic CO_2 emissions are prescribed from the extended SSP scenarios in Meinshausen et al. [2020] until the year 2500 CE, with zero emissions assumed afterwards. Anthropogenic CH_4 emissions are prescribed from Gidden et al. [2019] until 2100 CE and have been extended following Meinshausen et al. [2020] and Kleinen et al. [2021], with a linear decrease of non-agricultural emissions to zero by 2250 CE, while emissions from agricultural sources stay constantly elevated beyond this time. For further details on the emission scenarios refer to Fig. 1 and the Supplementary Information of Kaufhold et al. [2025]"

Reviewer Point P 2.10 — line 179: I'm confused as to the point of doing standalone EBM calculations covering the full span to year 4000. Doing this is by no means equivalent to really running a full simulation of the climate system with eg, for experiment EA the climate and ice sheets model of MPI-ESM but the EBM calculation of AWI-ESM. The evolution of SMB from the standalone EBM will likely quickly become quite different from what would have happened in the full system because it is so dependent on the changing ice sheet topography on the timescales in question, and the change in topography it sees are those that were produced by the original MPI-ESM EBM, not the AWI-ESM EBM that is being tested, surely? It makes sense that the analysis to be done just focuses on a timeslice very near the beginning of the simulation (line 180),

when the climate-sensitivity-dependent forcing signal has become different between models but the topography feedbacks have not yet become too significant - so perhaps you don't need to mention that the standalone EBM calculations were done for the later periods too?

Reply: We have removed the sentence in which we state that the simulations were run for the period from year 2000-4000.

Reviewer Point P 2.11 — line 186: "to the EBM used" would be better

Reply: We reformulated the sentence.

Reviewer Point P 2.12 — line 217: Many factors control the simulated rate of AMOC decline in a model and I don't feel like ocean resolution alone is a strong or reliable enough factor that it's worth singling it out.

Reply: We agree and have removed this.

Reviewer Point P 2.13 — line 220 and on: are the numbers reported the T and P natively from the climate model or after they have been downscaled by the EBM / ice sheet coupling?

Reply: We added the following sentence to make it clear that we use the native climate model grid: "All temperature and precipitation changes reported below are calculated on the respective models' native grids."

Reviewer Point P 2.14 — line 228: the CLIMBER-X SSP5-8.5 simulation invites a simplistic conclusion - there's an almost qualtitative diffrence in the sea-level contribution of the GrIS on a 1000yr timescale linked to AMOC resilience (and melt-elevation feedback), where in the SSP5-85 simulations here by year 300 you have lost 1/3 of the GrIS if the AMOC shuts down, or see complete deglaciation if it does not. Is that going too far?

Reply: We agree that the contrasting outcomes are noteworthy. However, since this behaviour is observed in only one model, we consider it premature to draw such a broad conclusion regarding the relationship between AMOC resilience and Greenland Ice Sheet retreat.

Reviewer Point P 2.15 — line 245: how do the spin-up icesheet extents compare to BedMachine for the 3 models? Could ice thickness differences (or surface height) compared to BedMachine also be plotted. Surface ice velocities compared to modern observations would also be useful to interpret whether the simulations of ice dynamics are likely to be in a reasonable ballpark. These features might help explain the different retreat patterns of the icesheet for each model, the differences between which are not discussed.

Reply: We thank the reviewer for this suggestion. We have added a figure to the Appendix comparing the simulated ice thickness in the three models with BedMachine observations, including the corresponding differences. We also discuss these differences in the revised manuscript.

Reviewer Point P 2.16 — Figure 2: printed at A4, the 15-panel map figures such as this are really too small to make out some of the geographical detail in each panel

Reply: We agree that some geographical detail is difficult to discern when the figure is printed at A4 size. However, we consider the current layout to be the best compromise between displaying spatial detail and enabling direct comparison across all models. We have also adopted a consistent layout across the figures to facilitate comparisons between the ice-sheet and atmospheric variables.

Reviewer Point P 2.17 — line 250: The lower end of the projection results here are all less than would be implied by a simple continuation of the currently observed 0.7mm/yr noted in the Introduction. Does this imply that the models in general do not reproduce currently observed mass loss rates? This deserves some comment. It might also be useful for some readers, since they generally look pretty linear, to quote the mass loss results as rates rather than just total amounts at different time horizons.

Reply: We have added this information to the revised results section.

Reviewer Point P 2.18 — line 256: I think the consistent difference in retreat patterns across the models interesting. In theory this aspect is the sort of thing that a MIP is well-placed to understand and exploit in terms of perhaps saying which models behave more realistically than others, but this feature is mentioned and then not returned to?

Reply: We have expanded the discussion to provide a more detailed interpretation of the sensitivity experiments and to explain the potential reasons for the distinct ice-sheet retreat patterns across the models. However, we refrain from assessing which retreat patterns are more realistic, as there is no objective basis for determining this.

Reviewer Point P 2.19 — line 284: whilst it conforms to expectation that surface melt would be found to be the dominant term in both SMB and the future GrIS mass budget generally, it is maybe worth noting that in all of the modelling systems here melt is the aspect to which most attention is paid in terms of process modelling and downscaling, and the aspect most amenable to lower resolution climate modelling of large-scale thermodynamics. That makes this conclusion a bit of circular argument, where other (more unlikely) factors such as fast ice dynamics, incursion of warm ocean water under the interior of the ice sheet where bedrock is below sea level and dynamics of atmospheric precipitation are barely represented, if present at all, in the class of climate models and ice sheet coupling used in this study.

Reply: We agree that the relatively coarse resolution of the coupled climate–ice-sheet models limits their ability to resolve certain processes, such as ice–ocean interactions within fjords and orographic precipitation. However, model resolution should also be considered in the context of the timescales investigated. The coupled models used in this study provide climate-ice sheet simulations over millennial timescales that is not readily achievable with higher-resolution modelling systems. Furthermore, the ice-sheet model resolutions of 5–16 km are comparable to those used in stand-alone ice-sheet projections, including those conducted within ISMIP6. While some processes may not be explicitly resolved, the models can nevertheless capture many aspects of rapid ice-sheet dynamics. For example, one of our model configurations is capable of simulating Heinrich events, which involve rapid changes in ice discharge. We therefore consider our results informative regarding the relative contributions of different processes to ice-sheet mass loss within the limitations of the modelling framework. Nevertheless, we acknowledge that the dominance of surface melt in our simulations should be interpreted in the context

of the processes represented in the models, rather than as evidence that other processes are unimportant for the future evolution of the GrIS.

Reviewer Point P 2.20 — Figure 3: Showing the information only as anomalies (and having defined calving mass loss as a positive-sense flux in equation 1) means it's really not obvious from the figure that the calving behaviour shown actually amounts to (line 77) reducing to zero across all models. Is the figure much less clear if absolute fluxes are shown?

Reply: Following the suggestion of Reviewer 1, we have clarified in the figure caption that the figure shows anomalies rather than absolute values. We prefer to show anomalies instead of absolute values as this ensures differences in the magnitude of changes are readily interpreted.

Reviewer Point P 2.21 — Presenting the information as area-integrals when (presumably) both the average flux and the area of the ice sheet are both changing simulataneously can also hide some interesting behaviour - especially as here when the vertical scale chosen makes the accumulation changes appear small. Separate plots of extent and area-average flux might be interesting, if you haven't looked at those?

Reply: We thank the reviewer for this helpful suggestion. The changes in ice-sheet area are relatively small (Fig. 1), particularly under the lower-emission scenarios that are the focus of our analysis. We therefore retain the current presentation. However, to acknowledge the potential influence of changes in ice-sheet area on the area-integrated SMB, we have added the following clarification to the manuscript: " Although changes in ice-sheet area also affect the integrated SMB, their influence is relatively small, particularly under lower-emission scenarios."

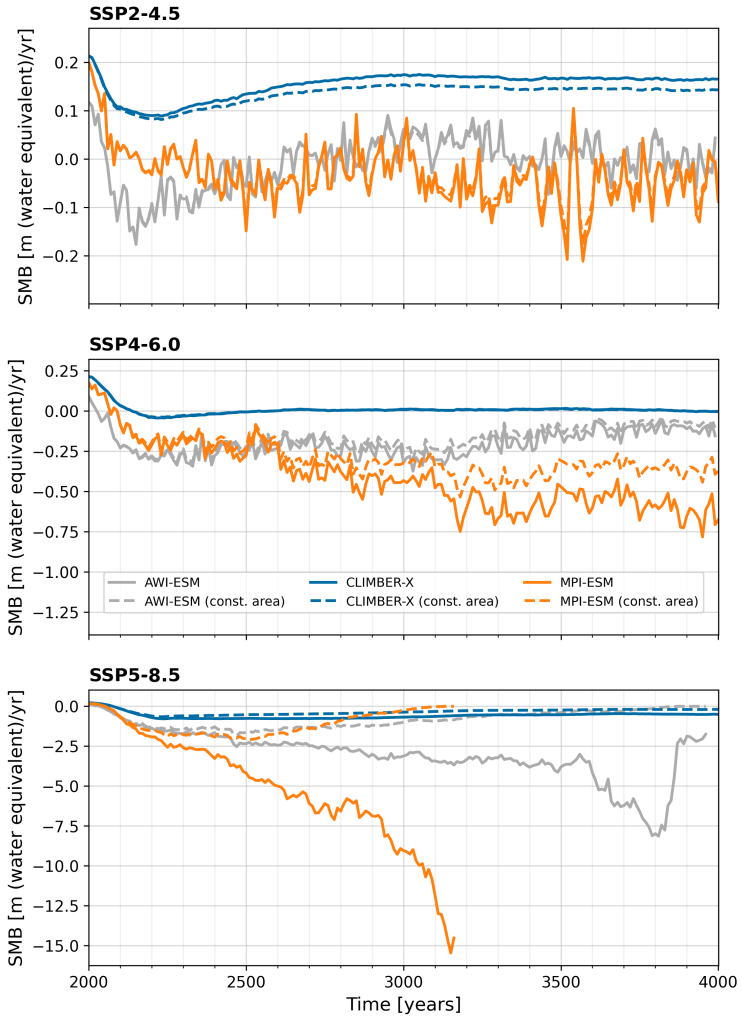

Figure 1: Mean surface mass balance for the three coupled models under the three emission scenarios. Dashed lines indicate the corresponding values calculated assuming a constant surface area equal to the initial ice-sheet area.

Reviewer Point P 2.22 — line 347: 3.5 felt like another section that would have benefited from more description of the results in question and walking the reader through from there to the author's conclusions. At the moment it feels a bit too condensed - the results, analysis and conclusion of all 9 of the factorial experiments trying to unravel the influence of each of the 3 main contributions selected take only one page of text. This section is where it might fit to attempt some explanation of why the main experiments with each model see the very different retreat patterns that they do. Might that lead to some insight into the model differences that could help suggest which of the projections might be more or less likely. The difference in rates of ice loss in the next few centuries in the SSP5-85 scenario between the MPI-ESM and CLIMBER-X simulations would imply very different strategies for sea-level adaption and any information to help reduce the uncertainty would be very valuable. The wording of the section title and second sentence could also be more precise - it is really just the influence of the initial surface topography that is beign isolated

in this experiment design. Feedbacks between the evolving topography and the climate and melt calculation are clearly very important for the long-term evolution, and those are not isolated here.

Reply: We have expanded the discussion to provide a more detailed interpretation of the sensitivity experiments and to explain the potential reasons for the distinct ice-sheet retreat patterns across the models.

Reviewer Point P 2.23 — line 358: I don't think I agree that that the topography is responsible for the least spread hypsometrically in figure 6(c,i), A6, A7. Although all three topographies have a similar ELA, below 1000m where melting has the greatest effect the magnitude of the ablation seems to vary more significantly than that produced by varying the EBM

Reply: We agree and now specifically mention this fact in the discussion. However, the integrated and structural differences in the SMB and its components are still smaller compared to the other two sensitivity categories.

Reviewer Point P 2.24 — line 393, 396: is it true to conclude for all projection timescales in these results that initial geometry plays minor role in these results, I suspect on shorter timescales it is a much more important factor (as noted in line 390 that cites Goelzer2018), and that should not be forgotten in these conclusions - different readers will have different timescales in mind as a priority. This is not the first study that has noted this sort of conclusion on the limited influence of initial conditions, eg Coulon et al. Nat Comms (2025) for Antarctica

Reply: We reformulated these sentences to make sure that we are referring to coupled climate-ice sheet simulations and millennial timescales. It reads: "Considering that data assimilation techniques typically only capture memory effects spanning the observational period and that models often exhibit substantial initial drifts, the finding that the initial ice-sheet geometry appears to be less critical is encouraging, particularly for coupled climate–ice sheet simulations covering millennial timescales. More specifically, our results suggest that, on millennial timescales, uncertainty arising from differences in initial ice-sheet geometry may not be the dominant source of uncertainty in sea-level rise projections from coupled climate–ice sheet models."

Reviewer Point P 2.25 — The sensitivity tests here can only be done before the evolving ice sheet topography starts to play a role. That being the case, I don't think the experiment design can rule out representation of ice model dynamics as a crucial factor in obtaining "reliable long-term projections of the GrIS" (Abstract). The conclusions written in this paper are only about the importance of the representation of climate and SMB compared to the initial condition which their analysis does support (and line 385 expresses well in terms on the initial value vs boundary-type problem), but I think this issue should be noted.

Reply: We have added to the discussion and the conclusions section the qualifier that we do not explicitly investigate the contribution of ice dynamics to put our results in the correct context.

Reviewer 3

Reviewer Point P 3.1 — This study investigates the primary sources of divergence in multi-millennial (up to 4000 CE) Greenland Ice Sheet (GrIS) projections across three coupled climate-ice sheet models under various emission scenarios. While projected sea-level contributions remain relatively uniform and modest across models by 2100 (0.03-0.11 m), trajectories diverge dramatically on multi-centennial to millennial scales, reaching up to 3.5 m by 2500 and complete disintegration under high emissions by 3000 CE. Using targeted sensitivity experiments that isolate individual model components, the authors demonstrate that differences in the simulated atmospheric climate are the dominant drivers of long-term mass-loss spread, surpassing uncertainties in the representation of SMB processes, and far outweighing the influence of initial ice-sheet geometry.

The manuscript is well written, clearly structured, and features high-quality, informative figures. The experimental framework is well designed to address the primary research questions. However, several aspects regarding scenario setup, model resolution, temporal scales, and comparison with existing literature would benefit from further clarification and discussion. I believe that all of the comments below are relatively straightforward to address.

Reply: We thank the reviewer for the overall positive assessment of our manuscript.

Reviewer Point P 3.2 — 1. Contextualizing Findings with Previous Ensemble Studies: The core finding, that differences in atmospheric forcing and SMB schemes dominate over other uncertainties, closely echoes the conclusions of earlier ensemble work (Holube et al., 2022, doi:10.5194/tc-16-315-2022). In particular, Figures 6a and 6b strongly support those earlier findings. It would be valuable to include an explicit discussion of how the findings from the fully coupled, advanced models used here complement the more comprehensive large ensemble and statistical variance partitioning of that earlier work.

Reply: We agree and have added a short discussion to put our results into this wider context.

Reviewer Point P 3.3 — 2. Ice-Ocean Interactions and Model Resolution: While the conclusions regarding SMB dominance are well supported by the experimental setup, I wonder whether the relative importance of ice-ocean interactions might be underestimated due to the coarse resolution of the ice sheet models. Resolving narrow Greenlandic fjords and dynamic calving margins realistically requires high spatial resolution, which is necessarily limited in multi-millennial coupled runs. A discussion addressing the implications of this low resolution and how it might impact the balance between ocean-driven dynamic loss and atmospheric SMB loss would strengthen the paper.

Reply: We agree that the coarse spatial resolution limits the representation of ice–ocean interactions and may affect the simulated contribution of ocean-driven dynamic ice loss relative to SMB changes. We have added a statement to the revised manuscript to acknowledge this limitation while clarifying that it reflects a necessary compromise to make simulations over millennial timescales computationally feasible. The revised text reads: "The representation of ice–ocean interactions, particularly processes affecting marine-terminating glaciers and ice shelves, involves simplifications due to the coarse spatial resolution of the ocean models. This is a necessary compromise to make simulations over millennial timescales computationally feasible."

Reviewer Point P 3.4 — 3. SSP Scenario Extension: The simulation design is well suited to address the core research questions. However, the manuscript needs greater clarity on how the standard SSP scenarios were extended beyond 2100/2300 to the year 4000 CE. In particular, please clarify the exact role that CLIMBER-X played in determining the atmospheric greenhouse gas concentrations across the models for this extended period.

Reply: We have added some more details on the extended emission-driven SSP scenarios: "As forcing, we utilized time-varying CO_2 and CH_4 levels computed in separate emission driven CLIMBER-X simulations for the different scenarios. Future anthropogenic CO_2 emissions are prescribed from the extended SSP scenarios in Meinshausen et al. [2020] until the year 2500 CE, with zero emissions assumed afterwards. Anthropogenic CH_4 emissions are prescribed from Gidden et al. [2019] until 2100 CE and have been extended following Meinshausen et al. [2020] and Kleinen et al. [2021], with a linear decrease of non-agricultural emissions to zero by 2250 CE, while emissions from agricultural sources stay constantly elevated beyond this time. For further details on the emission scenarios refer to Fig. 1 and the Supplementary Information of Kaufhold et al. [2025]"

Reviewer Point P 3.5 — 4. Timescales of SMB Forcing and Atmospheric Variability: The manuscript could provide a more explicit discussion of the relevant temporal scales governing SMB. This is highlighted by the use of dEBM, which relies on monthly mean input data while explicitly resolving the diurnal radiation cycle. This approach bridges two extremes but leaves out intermediate synoptic and weather-scale variability, which has been shown to strongly impact non-linear threshold processes and overall SMB (Zolles and Born, 2024, doi:10.5194/tc-18-4831-2024). In contrast, MPI-ESM utilizes hourly input forcing. Can the authors quantify or discuss how the choice of dEBM and the smoothing inherent to monthly averaging might affect the sensitivity experiments and calculated melt rates relative to higher-frequency forcing?

Reply: We agree that this is a point worth mentioning. We have added a paragraph to the discussion as follows: "In addition to differences in EBM formulation and tuning, the temporal resolution of the atmospheric forcing may influence the simulated SMB. For example, dEBM relies on monthly mean input data while explicitly resolving the diurnal radiation cycle, whereas MPI-ESM uses hourly atmospheric forcing. Consequently, dEBM does not capture synoptic and weather-scale variability in the atmospheric input, which can influence surface melt through nonlinear processes and threshold effects [Zolles and Born, 2024]. Temporal resolution is therefore another factor to consider when interpreting differences in SMB between model configurations, although its individual contribution cannot be isolated from the other differences in EBM formulation and forcing characteristics in our experiments."

Reviewer Point P 3.6 — 5. Importance of Initialization: The conclusion that initial ice-sheet topography (and initialization more broadly) plays a minor role in long-term divergence is only valid for the relatively narrow range of states tested here. Because all three models were initialized with very similar present-day configurations, the spread in initial geometry and thermal state is inherently limited. In a wider ensemble, or when comparing substantially different initialization techniques, initial conditions can introduce much larger uncertainties. The authors should explicitly caveat this finding and clarify that the conclusion reflects the specific, closely aligned initial states of the participating models rather than an exhaustive test of initialization uncertainty.

Reply: As also suggested by Reviewer 1, we have revised the manuscript to clarify that our conclusions regarding the influence of initial ice-sheet geometry apply to the specific model configurations considered

here. However, we do not agree that the initial ice-sheet states are very similar across our models. Unlike studies employing data assimilation to obtain initial states that closely match the present-day ice sheet, we follow a "come-as-you-are" spin-up approach, resulting in substantially different ice thickness distributions as well as ice-sheet extents. Therefore, we believe that our simulations encompass a broader range of initial states than would be expected from an initialization approach that explicitly constrains the ice-sheet geometry to observations, such as those commonly used in ISMIP6.

Reviewer Point P 3.7 — 6. Introduction and Scope of GrICMIP: The acronym and initiative GrICMIP is introduced without sufficient background context. It remains unclear to the reader whether GrICMIP represents a broader, formally coordinated community initiative with an established protocol spanning multiple planned publications or if it is simply a project-specific acronym coined for this particular study and model trio.

Reply: We agree that this was not clear enough in the initial manuscript. We added the following sentence to calrify this: "Here, we build upon these findings and present the results of the Greenland Ice sheet Coupled Model Intercomparison Project (GrICMIP), an activity that emerged from the German Climate Modeling Initiative PalMOD, which focused on developing coupled climate–ice sheet models for investigating long-term climate change."

Reviewer Point P 3.8 — 7. Structure and Readability: Although the overall writing quality is very high, several sections contain dense blocks of text that make following the detailed model descriptions difficult. Breaking up longer sections into more and shorter paragraphs would significantly improve readability and flow.

Reply: We agree and have broken up the dense blocks of text into smaller paragraphs.

References

G. Beobide-Arsuaga, T. Bayr, A. Reintges, and M. Latif. Uncertainty of ENSO-amplitude projections in CMIP5 and CMIP6 models. *Climate Dynamics*, 56:3875–3888, 2021. doi: 10.1007/s00382-021-05673-4.

N. E. Erickson and C. M. Patricola. Future projections of the El Niño–southern oscillation and tropical pacific mean state in CMIP6. *Journal of Geophysical Research: Atmospheres*, 128(21): e2022JD037563, 2023. doi: 10.1029/2022JD037563.

M. J. Gidden, K. Riahi, S. J. Smith, S. Fujimori, G. Luderer, E. Kriegler, D. P. Van Vuuren, M. Van Den Berg, L. Feng, D. Klein, K. Calvin, J. C. Doelman, S. Frank, O. Fricko, M. Harmsen, T. Hasegawa, P. Havlik, J. Hilaire, R. Hoesly, J. Horing, A. Popp, E. Stehfest, and K. Takahashi. Global emissions pathways under different socioeconomic scenarios for use in CMIP6: A dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development*, 12(4):1443–1475, 2019. ISSN 19919603. doi: 10.5194/gmd-12-1443-2019.

C. Kaufhold, M. Willeit, S. Talento, A. Ganopolski, and J. Rockström. Interplay between climate and carbon cycle feedbacks could substantially enhance future warming. *Environmental Research Letters*, 20(4):044027, Mar. 2025. ISSN 1748-9326. doi: 10.1088/1748-9326/adb6be. URL `http://dx.doi.org/10.1088/1748-9326/adb6be`.

T. Kleinen, S. Gromov, B. Steil, and V. Brovkin. Atmospheric methane underestimated in future climate projections. *Environmental Research Letters*, 16(9), 2021. ISSN 17489326. doi: 10.1088/1748-9326/ac1814.

K. D. Mankoff, X. Fettweis, P. L. Langen, M. Stendel, K. K. Kjeldsen, N. B. Karlsson, B. Noël, M. R. van den Broeke, A. Solgaard, W. Colgan, J. E. Box, S. B. Simonsen, M. D. King, A. P. Ahlstrøm, S. B. Andersen, and R. S. Fausto. Greenland ice sheet mass balance from 1840 through next week. *Earth System Science Data*, 13(10):5001–5025, Oct. 2021. ISSN 1866-3516. doi: 10.5194/essd-13-5001-2021. URL `http://dx.doi.org/10.5194/essd-13-5001-2021`.

M. Mastropierro, D. Peano, and D. Zanchettin. Drivers of and uncertainty in Amazon carbon sink long-term and interannual variability in CMIP6 models. *Biogeosciences*, 22:5231–5246, 2025. doi: 10.5194/bg-22-5231-2025.

M. Meinshausen, Z. R. Nicholls, J. Lewis, M. J. Gidden, E. Vogel, M. Freund, U. Beyerle, C. Gessner, A. Nauels, N. Bauer, et al. The shared socio-economic pathway (ssp) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, 13(8):3571–3605, 2020.

P. Testorf, C. Schannwell, M. Kapsch, and U. Mikolajewicz. Coupled climate-ice-sheet simulations reveal novel teleconnection between northern hemisphere ice sheets and the antarctic ice sheet. *Geophysical Research Letters*, 53(1), Jan. 2026. ISSN 1944-8007. doi: 10.1029/2025gl118959. URL `http://dx.doi.org/10.1029/2025GL118959`.

T. Zolles and A. Born. How does a change in climate variability impact the greenland ice sheet surface mass balance? *The Cryosphere*, 18(10):4831–4844, Oct. 2024. ISSN 1994-0424. doi: 10.5194/tc-18-4831-2024. URL `http://dx.doi.org/10.5194/tc-18-4831-2024`.