

Interactive comment on “CMIP5 model selection for ISMIP6 ice sheet model forcing: Greenland and Antarctica” by Alice Barthel et al.

Response to reviewers:

We thank the reviewers for their thoughtful reviews and are pleased that the manuscript was well received.

Both reviewers recommended to improve the conclusions section to highlight the main results of this study. We significantly increased the scope of the conclusions. They also both suggested minor corrections to improve the clarity of the manuscript, and we endeavored to implement these changes.

Reviewer #1 rightly pointed out the importance of the SAM and/or ASL in Antarctic regional variability. We thank them for noting that a comprehensive analysis of the SAM/ASL representation is beyond the scope of this study, but we agree that including a mention of this mode of variability adds to the manuscript. We therefore included it in the discussion section. Following their recommendation, we also fleshed out the appendices to summarize the findings from each of the appendix figures.

Reviewer #2 suggested changes to the variables notations, which we are happy to implement to improve the readability of the manuscript. They also shared concerns about referencing upcoming papers. We updated the references to (a) include the DOI of papers if applicable, (b) cite the publicly available project website if the relevant manuscripts are still unavailable.

Our responses to each of the reviewers comments are included in blue italic below.

Anonymous Referee #1

Summary This article summarizes the selection of atmosphere-ocean coupled climate models (AOGCMs) to use for forcing for the stand-alone ice sheet simulations as part of the CMIP6 ice sheet intercomparison project (ISMIP6). The manuscript summarizes the methods used to select the models and recommends 6 models each for use with Antarctica and Greenland ice sheets. The models used in the selection process are those from CMIP5 AOGCMs (CMIP6 were insufficiently available for testing at the time of this analysis). Three “core” models are chosen for both Antarctica and Greenland based on their fidelity to observations during the satellite record period (1979-2005). Three more models (“targeted”) were selected for use based on representation of a range of future atmosphere-ocean conditions from both the RCP2.6 and RCP8.5 emissions scenarios. This submission documents the selection criteria and subsequent of specific AOGCM selection for forcing of regional models for the ISMIP6. The work described is

new and unique in that it uses both atmospheric and oceanic observations (rather than just atmospheric as in previous work) in the selection criteria. Antarctica and Greenland are treated separately, and with some overlapping and some unique variables as part of the evaluation of the AOGCMs.

The manuscript is well written, represents a significant scientific advance w.r.t. model selection for boundary conditions for ice sheet models. I recommend it be accepted for publication in The Cryosphere with minor revisions and technical corrections as follows:

We thank Reviewer #1 for their thoughtful review and are pleased that the manuscript was well received. We thank Reviewer #1 for their suggested technical corrections and address them below.

Minor revisions: Much of the regional variability in Antarctica is related to the zonal asymmetry in the Southern Annular Mode (SAM; or likewise the depth, location, and seasonal migration of the Amundsen Sea Low, ASL). Some models do a better job than others at capturing this – which is different than the metrics of zonal jet location and strength. There are many atmospheric and oceanic metrics used to select the model criteria in this submission, although none directly measure whether or not the models capture asymmetric nature of the SAM (although the combination of oceanic and atmospheric metrics used may indeed capture it indirectly). A full analysis of this (whether or not models capture this asymmetry, not to mention how, exactly, to measure if the models do) is beyond the scope of this paper. I do feel, however, that some mention is worthwhile – do you believe your metrics indeed capture this even if indirectly? Or do you think some of the regional biases might be due to a particular model's lack of an ASL? A model's fidelity or lack thereof to ASL could help explain some of the regional discrepancies in projected changes as well. (e.g. M. Holland, L. Landrum, Y. Kostov and J. Marshall, 2016, Sensitivity of Antarctic sea ice to the Southern Annual Mode in coupled climate models, Clim. Dyn., DOI 10.1007/s00382-016-3424-9; J. T. M. Lenaerts, J. Fyke, B. Medley. The signature of ozone depletion in recent Antarctic pre- cipitation change: a study with the Community Earth System Model, 2018. Geophys. Res. Lett., 45, 23, <https://doi.org/10.1029/2018GL078608>)

We agree that the role of the SAM and/or ASL in regional variability is worth mentioning. We include this point in the discussion section of the final manuscript.

A couple sentences summarizing the figures/main point for each appendix would be C2 helpful (have one sentence for Appendix C, none for A, B).

We added text to the appendices to summarize the main points of each appendix.

Conclusions? Please finish!

We added text to flesh out the conclusions and highlight the goals and main findings of the study.

Technical corrections (I can't figure out how to cut and paste Greek chi here so I write [chi]): Line 147: "historical metrics [chi] described above" but chi is not defined above. I believe [chi] in this case is the RMSE from the observations for each given variable – state this

We modified this line to remove ambiguity: the historical metrics are described above, although chi is only used below.

Lines 315-322: Section 4.3 Top 3 (Greenland) Last sentence in first paragraph ("model 1, model2") sounds like a placemaker – eliminate or re-write

The placemaker "model1, model2" is intentional here. It is to highlight that the MIROC5 model was strategically chosen (and imposed), while the other two models were selected among the

ensemble through our selection algorithm. The selection of model1 and model2 is explained in the following sentence.

Lines 435-455 Check figure numbers. Mismatch between titles (in bold) and descriptions below (e.g. lines 439: “C2 Robustness of Antarctic . . .” Followed in line 440 by “Table C3 lists the . . .”

We agree that the numbering of the appendices was confusing (section C1 applied to Tables C1 and C2, section C2 applied to Tables C3 and C4...). We adjusted the numbering to avoid confusion: the sections in appendix C are unnumbered, while the tables are numbered as C.1, C.2, etc. to be consistent with the appendix figure numbering system.

Figure 1. Regional oceanic boundaries (and some of the text over the map of the continent) for Antarctica are difficult to see (very difficult in the printed version – better on the screen) – recommend trying for different colors, or perhaps thicker outlines of the regions. The most difficult regional texts are “Weddell (WS)” followed by “Amery (AM) caption: “Greenland . . . inside the usual boundaries of MAR simulations” define MAR?

We thank Reviewer #1 for their feedback. We changed the color of the continental mask to improve the readability of Figure 1 in the revised manuscript. We adjusted the caption to avoid confusion.

Figures 2, 3, 5, 6, A.1, A.2, B.1, B.2 The symbols denoting models that were in top3 and top6 ensembles are very difficult to see (and not stated in captions for A.1, A.2). Figures 3 and 6 highlight w different colors so perhaps not as important in these, however in the other figures these symbols need to be easier to spot – with color, or bold, or?

We added color highlights to these figures in the revised manuscript.

Table C2. Rewrite caption . . . says “three top models” and give statistics for four models (which are the four that give the two top-three combos) . . .

We agree that the wording of the Table title was confusing. We adjusted the title to reflect that the model considered are those included in the possible combinations making the top 3.

Anonymous Referee #2

With the aim of selecting a set of global climate models that represent best the current and projected climate of the Greenland (GrIS) and Antarctic ice sheets (AIS) to force the ice sheet models of ISMIP6, the authors evaluate, compare and rank 33 CMIP5 AOGCMs using observational data (present-day) combined with various atmospheric and oceanic metrics (scenarios). As a result, an ensemble of six AOGCMs (i.e. three core and three targeted) is selected separately for the GrIS and AIS. These models show the best agreement with present-day observations while maximizing the diversity of future projections. The authors show that CMIP5 models performing the best differ in Greenland and Antarctica, and that they do not represent atmospheric and oceanic processes equally well.

This is a sound, very well written study that is highly relevant for the Cryosphere community. The AOGCMs selected here will be used to force ice sheet models participating in the ISMIP6 project. Using the outputs of the best performing AOGCMs as forcing will prove essential to better project the mass balance of the GrIS and AIS in a future warming climate, and to improve

estimates of their relative contribution to global sea level rise. I deem that the manuscript should be accepted for publication in the Cryosphere after applying some **minor revisions**. The authors can find my comments hereunder.

We thank Reviewer #2 for their thoughtful review and are pleased that the manuscript was well received. We thank Reviewer #2 for their comments and suggested technical corrections, which we address below.

General comments:

1) The conclusion section should be reformulated to stress the main results of the study, i.e. purpose of the inter-comparison exercise, which climate models have been selected to force the ice sheet models, some perspective and future work based on e.g. CMIP6 models. The current conclusion section should better be moved to the discussion section. In addition, reference to Tables 2 and 3 appear for the first time in the conclusion section, while they should better be discussed at the end of Section 3.3 (Table 2) and Section 4.3 (Table 3).

Following the reviewers' suggestion, we added text to flesh out the conclusions and highlight the goals and main findings of the study. We also refer to Table 2 and 3 in Sections 3.3 and 4.3, as suggested.

2) The authors refer multiple times to forthcoming papers that are currently in preparation. I would strongly advise to remove those references or better use a personal communication statement as at e.g. **L61, L87, L144, L254-255**.

We had hoped for these papers to be published, or under review (with a DOI) by the time of publication of this manuscript. If it looks unlikely by the time of the final revision, we will replace these references with the link to the existing Wiki and/or personal communication statements.

3) The authors should define the acronyms (e.g. ta850, prw, ...) used for the evaluation metrics. These are currently not listed in the main manuscript making the interpretation of Figs. 2, 4, 5, 7 and A1 difficult. This should be done at **L107-111** (AIS) and **L113-114** (GrIS). For clarity, sea surface temperature in summer and sea ice extent in winter could be better defined as sst[s] (instead of tos[s]) and sie[w] (instead of mwsie). For consistency, I also suggest to replace $\partial prw[a]$ by $\Delta prw[a]$ in the main text and figures (e.g. L164-168). In addition, at L165-166, the authors refer to winter sea ice concentration (i.e. fraction of a pixel covered by sea ice) as opposed to sea ice extent (i.e. integrated area of pixels with a sea ice fraction > 0.15). Please clarify which quantity is used in both cases.

We adjusted the acronyms to improve readability, as suggested by the reviewer.

For $\partial prw[a]$, we do not replace it by Δ as the ∂ notation indicates a different projected change: ∂ is for the difference divided by the mean over historical period. The text was adjusted to clarify the meaning of this metric. We thank the reviewer for pointing it out.

With regard to sea ice concentration vs. sea ice extent, it was indeed a mistake, we only used sea-ice extent. We corrected it in the text.

Point comments: L19: The authors should also refer to more recent studies such as Mougintot et al. (2019; GrIS) and Shepherd et al. (2018; AIS). See also additional references. *We included these references as suggested.*

L39: Remove “AOGCM” since it is first defined at **L41**. *Adjusted as suggested.*

L49: Could the authors provide a reference here (i.e. after ice shelves)? *We added a relevant reference in the revised manuscript.*

L50: Could the authors provide a reference? *We added a relevant reference in the revised manuscript.*

L53-54: The authors should add “e.g.” before “Noël et al., 2018” and “van Wessem et al., 2018”. *Adjusted as suggested.* For instance, Langen et al. (2017) and Niwano et al. (2018) also show good agreement between HIRHAM5 and NHM-SMAP RCMs and in situ measurements over the GrIS.

L107-109: I strongly suggest: “850 hPa (ta850; average of [...] precipitable water (prw), [...] pressure (psl), temperature (sst[s]) and winter sea ice extent (sie[w]) [...] jet strength (Jstr) and position (Jpos), [...] maximum in annual mean 850 hPa zonal wind [...]”. *Adjusted as suggested.*

L113-115: I strongly suggest: “[...] 700 hPa (ta700; average [...] at 500 hPa (zg500), inside the Modèle Atmosphérique Régional (MAR; Fettweis et al., 2017) [...]”. *Adjusted as suggested.*

L115: “do not significantly impact MAR results”. *Adjusted as recommended.*

L131: “ORCA025”. *Adjusted as recommended.*

L138: Could the authors provide a reference here? *We added a relevant reference in the revised manuscript.*

L141-142: “World Ocean Atlas (WOA; Locarnini and [...] 2018 WOA data (Locarnini [...])” *Adjusted as suggested.*

L166 and L169: Add “(ΔT)” after “ocean temperature”. *Adjusted as suggested.*

L173: At L156, the authors refer to 7 metrics for Greenland, while “6” is stated at L173. Do the authors discard $\Delta zg500$ from the comparison between future climate projections? Please, clarify. *$\Delta zg500$ is discarded from the future projections as a change in gridpoint geopotential height is not immediately meaningful as an indicator of a change in atmospheric circulation, as it can be due to a strengthening or a spatial shift in patterns. Thus, we prefer to include the jet metrics for Antarctica, but have not equivalent metric in Greenland.*

L182: Add “(Fig. 2a)” after 0.13 and “a” after “Figure 2”. *Adjusted as suggested.*

L183: Add “(blue)”, “(brown)” and “(yellow)” after “sub-surface ocean”, “atmosphere” and “surface ocean”. Same at **L240:** “(pink)”, “(red)” and “(light blue)”; at **L249** “(yellow)”, **L254:** “(brown)” and “(dark blue)”. *Adjusted as suggested.*

L208: I suggest: “We highlight the 3 core (red) and 3 targeted (yellow) AOGCMs selected in [...]” and remove **L218-220:** “In Fig. 3b [...] Amundsen region”. *Adjusted as suggested.*

L236: Top 3 (core models). **L248:** Top 6 (targeted models) and same at **L314** and **L328**. *We adjusted the title as follows:*

L242: Add "(dashed)" after "median". *Adjusted as suggested.*

L250: Do the authors mean "showing similar median projections under RCP8.5"? Please, clarify. *We reformulated the text to clarify the meaning.*

L266: The authors certainly mean "(Sections 4.1 and 4.2) and ensemble selection (Section 4.3)".

L303: "highlighted in Fig. 6b". *Adjusted as suggested.*

L307: $R^2 = 0.31$ is a weak correlation. Please, clarify. *A correlation with $R^2 = 0.31$ is still a moderate correlation ($R > 50\%$), while we would consider $R^2 \leq 0.25$ to be a weak correlation.*

L312: "[...]" show that RCMs outperform global climate models "[...]". *Adjusted as suggested.*

L315: What about EC-EARTH? No values are shown in e.g. Fig. 6. Could the authors elaborate? *At the time of the analysis, EC-EARTH future projections were not available for ocean data. We adjusted the wording to reflect that EC-EARTH was also disqualified due to data unavailability.*

L317: Add "Fig. 7a" after "median". *Adjusted as suggested.*

Section 4.3.1: The authors should refer to Figs. C3 and C4 here. *Adjusted as suggested.*

L333: For consistency, ΔT BB (instead of Baffin Bay) and "ACCESS1-3". *Adjusted as suggested.*

L347: Do the authors mean that a similar evaluation/model selection and ranking is not planned/possible using CMIP6 models? Or that the evaluation/selection of CMIP6 models was not performed in the current study? Please, clarify. *We reformulated the text to clarify the meaning.*

L408: "ACCESS1-3". *Adjusted as suggested.*

Stylistic comments:

We adjusted the stylistic comments as suggested. We thank Reviewer #2 for their detailed reading, and their effort to improve the readability of the manuscript.

L12: Maybe "limitations" instead of "constraints". **L17:** I suggest: "[...] most uncertain contributors to global sea-level rise over multidecadal to millennial timescales.". **L35:** I suggest: "[...] and oceanic forcing contribution to the mass balance of both ice sheets vary greatly, and depend on [...]". **L44:** Maybe "converting" instead of "translating". **L45-46:** I suggest: "[...] resolution that is too coarse [...] gradients impacting the surface climate of the ice sheets [...]". **L55:** Maybe "unable" instead of "challenged". **L73:** I suggest: "[...] some of the limitations of the selection procedure, and discuss [...]". **L111:** "[...] (in $m s^{-1}$), compared to time-slice [...]". **L203:** Maybe add "(core)" after "top 3 models". **L206:** "multi-model". **L225:** "but this region is projected to warm moderately [...]". **L235:** Remove "to choose from". **L238:** I suggest: "The correction is robust and removes [...] a time and changes the weight [...]". **L257:** Maybe "large number" instead of "high number". **L299:** Remove one of the double "an". **L359:** I suggest: "key processes for projections may still be missing.". **L366:** "models are assessed". **L367:** "evidenced in our analysis.". **L369:** "Concerning independence". **L387:** "e.g. Agosta et al., 2015" and "Meijers et al., 2012" before "Sallée et al., 2013". **L392:** "their results differ from the current study [...] ocean-driven basal melt". **L397:** "the different model performance". **L401:** "ice flux of the different ice sheets". **L403:** Maybe "reasonable" instead of "feasible". **L404:** "RCP scenarios, ..., parameters setting, ...". **L415:** I suggest: "better or project climate warming at different rates.". **L423:** "We refer readers interested in the [...] simulations to Slater et al. (2019)."

Figures:

Fig. 1: Does the grey mask in Fig. 1a also represent the AIS regions above 2000 m a.s.l. as in Fig. 4c? Please clarify in the caption.

No, the grey mask is over land for Antarctica, as stated in the caption (“For Antarctic atmosphere and surface ocean metrics, we considered the domain south of 40°S over ocean (color shading)”).

L107: Does the blue rectangle in Figs. 1a,c represent the integration domain of MAR? If so, “standard lateral boundaries of MAR (REFs for AIS and GrIS)”. **L112:** State the time period used for the “reference historical climatology”. In addition, move the titles of Figs. 1a and c upward so that they do not overlap with the figures. **Fig. 2:** For consistency replace legend items “surf. bias” and “ocean bias” by “surface ocean” and “sub-surface ocean”.

Adjusted as suggested.

What do the vertical bars in Fig. 2a represent? Please, clarify.

The (light gray) vertical lines were added to increase the readability the figure: i.e. they link the various metrics of a given model so that we can easily compare the biases of one model.

In Fig. 2b, add “ Δ ” before DML, Amery, Totten, ... The authors should also explicitly state that the horizontal dashed line represents the median of the models. Add “(core)” after top3 and “(targeted)” after top6.

In this manuscript, we reserve “ Δ ” as an indicator of the difference between end of 21st C and end of 20th century conditions. To avoid confusion, we do not use Δ for the historical bias metrics. We adjusted the other points as suggested.

Fig. 5: Same comments as for Fig. 2. The blue legend item in Fig. 5a should be “sub-surface ocean”. In Fig. 5b, add “ Δ ” before SPG, Baffin Bay ... In Fig. 5c, “zg500” instead of “zg550hPa”.

We thank the reviewer for noticing these details, we adjusted most of them as suggested. We decided to reserve the use of “ Δ ” to future changes (see comment above).

Fig. 4: For better contrast, I strongly suggest using a red line instead of the purple one for MIROC-ESM-CHEM. **Fig. 7:** Use a red line instead of the purple one for MIROC5.

We thank the reviewer for this suggestion. We adjusted the color of the pink line to increase contrast, as we prefer to avoid using red (already used for the “top 3” selection in other figures).

Fig. 6: The figure titles and labels are too small and almost unreadable. Please, enlarge.

Adjusted as suggested.

Figs. B1 and B2: As for Figs. 3 and 6, highlight the 3 core models in red and the 3 targeted models in yellow. *We will adjust the colors in the final manuscript.*

Additional references: 1) Shepherd et al. (2018):

<https://www.nature.com/articles/s41586-018-0179-y> 2) Mougnot et al. (2019):

<https://www.pnas.org/content/116/19/9239> 3) Langen et al. (2017):

<https://www.frontiersin.org/articles/10.3389/feart.2016.00110/full> 4) Niwano et al.

(2018): <https://www.the-cryosphere.net/12/635/2018/>

Other changes following C. Kittel comments

p4, L85; P12,L321: more 6-hourly variables are needed for forcing an atmospheric regional model (air temperature, humidity and surface pressure are missing). When I looked to the CMIP5 outputs last year, nearly all models having the 6-hourly winds had the other ones but it wasn't true for all the models and specifying the other variables could be more correct. *Corrected*

p4, L106: Gossart et al., 2019 do not say that ERA-Interim is the best reanalysis. Authors rather claim that it is ERA5. *Citation removed*

Figure 1, p5: The labels (A and C) are superimposed on the figure. *Corrected*

P12,L324. I'd suggest to replace MAR by RCM and climate models by GCM otherwise the sentence also suggests that MAR is better than RACMO (another climate model) Which is in fact true for the SMB but it is not mentioned in Noel et al., 2018. *Corrected*

CMIP5 model selection for ISMIP6 ice sheet model forcing: Greenland and Antarctica

Alice Barthel¹, Cécile Agosta², Christopher M. Little³, Tore Hattermann^{4,5}, Nicolas C. Jourdain⁶, Heiko Goelzer^{7,8}, Sophie Nowicki⁹, Helene Seroussi¹⁰, Fiammetta Straneo¹¹, and Thomas J. Bracegirdle¹²

¹Los Alamos National Laboratory, Los Alamos, NM, USA

²Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91198 Gif-sur-Yvette, France

³Atmospheric and Environmental Research, Inc., Lexington, Massachusetts, USA

⁴Norwegian Polar Institute, Tromsø, Norway

⁵Energy and Climate Group, Department of Physics and Technology, The Arctic University – University of Tromsø, Norway

⁶Univ. Grenoble Alpes/CNRS/IRD/G-INP, IGE, Grenoble, France

⁷Institute for Marine and Atmospheric research Utrecht, Utrecht University, Utrecht, the Netherlands

⁸Laboratoire de Glaciologie, Université Libre de Bruxelles, Brussels, Belgium

⁹NASA GSFC, Cryospheric Sciences Branch, Greenbelt, USA

¹⁰Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

¹¹Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

¹²British Antarctic Survey, Cambridge, UK

Correspondence: Alice Barthel (abarthel@lanl.gov)

Abstract. The ice sheet model intercomparison project for CMIP6 (ISMIP6) effort brings together the ice sheet and climate modeling communities to gain understanding of the ice sheet contribution to sea level rise. ISMIP6 conducts standalone ice sheet experiments that use space- and time-varying forcing derived from atmosphere-ocean coupled global climate models (AOGCMs) to reflect plausible trajectories for climate projections. The goal of this study is to recommend a sub-set of CMIP5 AOGCMs (3 core + 3 targeted) to produce forcing for ISMIP6 stand-alone ice sheet simulations, based on: i) their representation of current climate near Antarctica and Greenland relative to observations, and (ii) their ability to sample a diversity of projected atmosphere and ocean changes over the 21st century. The selection is performed separately for Greenland and Antarctica. Model evaluation over the historical period focuses on variables used to generate ice sheet forcing. For stage (i), we combine metrics of atmosphere and surface ocean state (annual- and seasonal-mean variables over large spatial domains) with metrics of time-mean sub-surface ocean temperature biases averaged over sectors of the continental shelf. For stage (ii), we maximize the diversity of climate projections among the best performing models. Model selection is also constrained by technical [constraints](#)[limitations](#), such as availability of required data from RCP2.6 and RCP8.5 projections. The selected top 3 CMIP5 climate models are CCSM4, MIROC-ESM-CHEM, and NorESM1-M for Antarctica, and HadGEM2-ES, MIROC5 and NorESM1-M for Greenland. This model selection was designed specifically for ISMIP6, but can be adapted for other applications.

1 Introduction and objectives

The Greenland and Antarctic ice sheets represent the largest and most uncertain ~~contribution to~~ contributions to global sea-level rise over multidecadal to millennial ~~timescale~~ sea-level-rise timescales. During the last three decades, satellite observation captured rapid mass loss from both ice sheets (~~Khan et al., 2014; Mougnot et al., 2014; Zwally et al., 2011; Velicogna, 2009~~) (Velicogna, 2009; Zwally et al., 2011; Khan et al., 2014; Mougnot et al., 2014; Shepherd et al., 2018; Mougnot et al., 2019). Both atmospheric and oceanic changes have been identified as drivers of observed mass loss, although regional mechanisms vary. For example, rising air temperatures over Greenland lead to increased surface melt causing direct mass loss (Trusel et al., 2018; Fettweis et al., 2017). Enhanced surface melt water production also destabilizes the margins of the ice sheet (van den Broeke, 2005; Banwell et al., 2013) and lubricates the ice flow at the bed (~~Kendrick et al., 2018; Andrews et al., 2014~~) (Andrews et al., 2014; Kendrick et al., 2018). Ocean interactions with the ice sheet occur in Greenland fjords, where a combination of on-shore ocean heat transport, estuarine-type circulation, subglacial meltwater runoff, and calving processes influence glacier terminus position and ice discharge (Straneo and Cenedese, 2015). In Antarctica, most of the ice sheet's mass loss is mediated through floating ice shelves. Melting at the ice shelf underside, which affects ice flow dynamics, is mainly controlled by the extent to which ocean dynamics along the continental margin allow intrusion of offshore ocean heat into the ice-shelf cavities, leading to distinct regimes operating in 'warm' vs. 'cold' continental shelf regions (e.g., Dinniman et al., 2016; Thompson et al., 2018). Rising air temperatures and associated surface melting are thought to be responsible for the collapse of ice shelves around the Antarctic Peninsula (Domack et al., 2005) and subsequent speed up of grounded ice flow (Rignot et al., 2004), while surface melting is currently limited in most other parts of the continent (e.g., Trusel et al., 2013). In the future, increased water vapor transport in a warmer atmosphere may lead to increased surface accumulation in Antarctica (Frieler et al., 2015; Palerme et al., 2017) together with increased melting over Greenland (Franco et al., 2013) and the Antarctic ice shelves (Trusel et al., 2015). Besides this general pattern, the spatial distribution and magnitudes of atmospheric and oceanic ~~mass balance contributions vary greatly over~~ contributions to the mass balance of both ice sheets vary greatly, and depend on synoptic-scale climate variability and physical processes at regional and smaller scales.

The ice sheet model intercomparison project for CMIP6 (ISMIP6) effort brings together the ice sheet and climate modeling communities to gain understanding of the ice sheet contribution to sea level rise (Nowicki et al., 2016). (~~Due to the delay in the~~ CMIP6 AOGCM dataset release, ISMIP6 revised the protocol described in Nowicki et al. (2016) to utilize climate forcing from the CMIP5 dataset (~~Nowicki et al., in prep.~~) (Nowicki, 2019). ISMIP6 conducts standalone ice sheet experiments that use space- and time-varying forcing derived from atmosphere-ocean coupled global climate models (AOGCMs) to reflect plausible trajectories for climate projections, building on earlier coordinated experiments which applied ad-hoc boundary conditions either constant in time, or imposed as an abrupt perturbation (Pattyn et al., 2013; Bindshadler et al., 2013; Levermann et al., 2014). However, this effort requires ~~translating~~ converting AOGCM output to forcing for ice sheet models, posing several challenges. First, climate models from the Coupled Model Intercomparison Project (CMIP) have a horizontal resolution that is

too coarse to accurately represent sharp ice-sheet topographic gradients ~~that impact surface climate over~~ impacting the surface
50 climate of the ice sheet (e.g. melt, wind, precipitation). Ocean components cannot represent narrow fjords connecting the deep
ocean and tidewater glaciers around Greenland (Straneo et al., 2012), or the ocean eddies involved in poleward heat transport
across continental shelves (Stewart et al., 2018), or ocean circulation beneath ice shelves (Asay-Davis et al., 2017). Second,
AOGCMs poorly represent polar-specific processes that have a major impact on the ice sheet surface climate (e.g. snowpack
evolution, cloud and boundary-layer processes) (Favier et al., 2017).

55 These limitations can be addressed by using regional climate models adapted for the polar regions. On the atmosphere side,
polar-oriented regional climate models (RCMs) have proved to provide more realistic surface climate than direct AOGCM
outputs for both the Greenland ice sheet (Noël et al., 2018; Fettweis et al., 2013) (e.g., Noël et al., 2018; Fettweis et al., 2013)
and the Antarctic ice sheet (van Wesseem et al., 2018; Agosta et al., 2019) (e.g., van Wesseem et al., 2018; Agosta et al., 2019).
On the ocean side, a number of models have recently added the capability to represent ice shelf cavities and ice/ocean inter-
60 actions (e.g., Dinniman et al., 2016). However, ocean simulations are still ~~challenged~~ unable to provide non-biased solutions
from a pan-ice sheet perspective, and remain computationally expensive, which probably explains the small number of existing
projections of ice-shelf basal melting (Timmermann and Goeller, 2017; Naughten et al., 2018). Thus, the ISMIP6 steering
committee has proposed the following strategy to convert AOGCM outputs into ice sheet forcing: surface forcing is provided
by AOGCMs dynamically downscaled with a polar-oriented atmospheric RCM (Fettweis et al., 2017), while ocean forcing
65 is computed by interpolating AOGCMs ocean temperature onto the continental shelf and by parameterizing ice shelf melt or
retreat rates, as detailed in Slater et al. (2019) and ~~Jourdain et al. (in prep)~~ Nowicki (2019).

The goal of this study is to recommend a sub-set of CMIP AOGCMs to produce forcing for ISMIP6 standalone ice sheet
simulations. This ensemble of AOGCMs aims to capture (i) plausible climate near Antarctica and Greenland over the historical
period, and (ii) a diversity of atmosphere and ocean warming rates over the 21st century. For evaluating AOGCMs we focus on
70 variables that are inputs of the downscaling methods defined to generate ice sheet forcing. Although it is technically possible
to select different AOGCMs for atmosphere and ocean forcing, we choose to use the same climate models across both realms
due to their inter-dependence in projections (e.g., Krinner et al., 2014; Bracegirdle et al., 2018). We thus perform a combined
assessment of both the atmosphere and ocean component of AOGCMs.

This paper describes the process utilized to select 6 AOGCMs to provide forcing for each ice sheet. This evaluation combines
75 observational/reanalysis data, metrics from existing studies and data produced specifically for this study. The methodology
to combine distinct metrics for the ocean and atmosphere into a single ranking is detailed in Section 2. The models are
selected independently for the Antarctic (Section 3) and the Greenland (Section 4) ice sheets. Finally, we present some of
the ~~assumptions underlying our analysis~~ limitations of the selection procedure, and discuss perspectives for future research in
Section 5.

2.1 General methodology

We analyze monthly output from 33 climate models of the CMIP5 ensemble, listed in Table 1. The ISMIP6 standalone experiment requires 3 coupled climate models to derive forcing fields for their core experiments (core), plus 3 additional models to extend the ensemble to a total of 6 models (targeted). To select the models, we first rank them according to their performance in reproducing observations over the 1979-2005 historical period (historical metrics, defined in Section 2.2). In a second step, we define climate change metrics over the 21st century (21C) under the RCP8.5 scenario (Section 2.3.1) in order to select a set of models that represents a diversity of 21C changes (Section 2.3.2). This two-step process is performed independently for the Antarctic and Greenland ice sheets.

The top 3 (core) models are those maximizing the diversity of climate change (Section 2.3.2, $n = 3$) among those fitting the following criteria:

1. the model must provide 6-hourly wind ~~outputs,~~ temperature and humidity, to be able to run an atmospheric regional climate model (18 models);
2. the model output must include required data fields under both the RCP2.6 and RCP8.5 scenario projections, following the revised ISMIP6 protocol (~~Nowicki et al., in prep~~) (Nowicki, 2019) (25 models);
3. the model must rank in the top half of the 33-model ensemble with regard to the historical metrics defined in Section 2.2 (17 models, Fig. 2a and Fig. 5a);
4. the model must not have any single climate change metric defined in Section 2.3.1 above 2 interquartile range (IQR, equal to quantile 75 % minus quantile 25 %) from the multi-model median projection (Fig. 4a and Fig. 7a).

For the additional 3 models (targeted), criteria used for the top 3 are relaxed, now including models without sub-daily frequencies for Antarctica, and including models with projected 21C changes above 2 IQR of the multi-model median. The models are selected to maximize the diversity of climate change across the top 6-model ensemble (Section 2.3.2, $n = 6$). As the selection method maximizing diversity tends to favor models with extreme values, we impose one model (within the top 6) which features 21C climate changes in the median range of the ensemble.

2.2 Historical metrics**2.2.1 Atmosphere and surface ocean metrics**

For the atmosphere and surface ocean, we consider variables that have an impact on RCM-modeled surface mass balance and for which reanalyses are reliable, following Agosta et al. (2015). All model outputs are bi-linearly interpolated onto a common regular longitude–latitude grid ($1.5^\circ \times 1.5^\circ$). For each variable that retains spatial information (described in the following paragraph), we calculate the spatial root mean square error (RMSE) for annual- or seasonal-mean values over 1980-

110 2004 (25 years). We take the European Centre for Medium-Range Weather Forecasts “Interim” re-analysis (ERA-Interim, 1979–present; Dee et al. (2011)) as a reference, since differences between reanalyses are much smaller than climate model biases (Agosta et al., 2015) and ERA-Interim was assessed to be the most reliable contemporary global reanalysis over Antarctica ([Bromwich et al., 2011](#); [Bracegirdle and Marshall, 2012](#); [Huai et al., 2019](#); [Gossart et al., 2019](#)) ([Bromwich et al., 2011](#); [Bracegirdle and Marshall, 2012](#)).

115 For Antarctica, we evaluate air temperature at 850 hPa ([ta850](#); average of summer and winter RMSE), annual precipitable water ([prw](#)), and annual sea-level pressure ([psl](#)), together with summer sea surface temperature ([sst\[s\]](#)) and winter sea ice extent ([sie\[w\]](#)), for the domain extending south of 40°S over the ocean (Fig. 1a). In addition to spatially-resolved variables, we include a metric of the historical CMIP5 vs ERA-Interim bias in westerly jet strength ([Jstr](#)), calculated as the maximum in annual mean zonal mean 850 hPa zonal wind between 10°S and 75°S (in m s^{-1}), ~~with comparisons made over compared to~~
120 time-slice means of the overlapping 1979-2005 period, as in [Bracegirdle et al. \(2018\)](#).

For Greenland, we evaluate air temperature at 700 hPa ([ta700](#); average of summer and winter RMSE), annual precipitable water ([prw](#)), and annual geopotential height at 500 hPa ([zg500](#)), inside the ~~RCM domain~~ domain of the “[Modèle Atmosphérique Régional](#)” (MAR; [Fettweis et al. \(2017\)](#)) and where the Greenland ice sheet is below 2000 m a.s.l. (bright shaded color in Fig. 1c). In this small domain, sea surface conditions do not ~~impact RCM~~ significantly impact MAR results
125 (Noël et al., 2014).

2.2.2 Sub-surface ocean metrics

The ISMIP6 standalone ice sheet oceanic forcing is derived from “far-field” salinity and potential temperature ([Slater et al., 2019](#); [Jourdain, in prep](#)). Consistent with this approach, our evaluation of sub-surface ocean properties is performed on regionally-averaged CMIP5 temperatures. Since the oceans around Greenland and Antarctica are characterized by different
130 geographic and dynamic regimes in observations (e.g., [Straneo et al., 2012](#); [Schmidtke et al., 2014](#); [Thompson et al., 2018](#)) and models ([Yin et al., 2011](#); [Little and Urban, 2016](#); [Levermann et al., 2014](#)), individual metrics are obtained for several sub-regions surrounding both ice sheets (Fig. 1b,d).

For this purpose, 1989-2009 time-mean ocean temperatures from each CMIP5 model are interpolated onto a common tripolar ORCA025 grid ([Ferry et al., 2012](#)), which has a quasi-isotropic resolution corresponding to 0.25 degrees in latitude, and 75 vertical layers with a thickness ranging from 1 m at the surface to 200 m at the bottom. We use a conservative
135 3d interpolation; if some parts of the ORCA025 grid are not covered by the CMIP grid, we extrapolate from the closest neighbor (horizontally above sills, then vertically to fill troughs behind sills). The regridding tools are made available on https://github.com/nicojourdain/SCRIPTS_CMIP5_ANOM_NOW (last access: 29 July 2019, [Dutheil et al., 2019](#)). Regionally averaged coastal ocean temperatures are then computed in six sectors around the Antarctic continent (Fig. 1b), which capture
140 different continental shelf and melting regimes. A maximum bottom depth criterion of 1500 m is used, together with an explicit limit for the northern boundaries in the large embayments in the Ross and Weddell Seas, to select ~~ORCA25~~ ORCA025 ocean cells that are located on the continental shelf near the coast. For Greenland, the ocean has been separated in four connected regions based on the major hydrographic regimes surrounding the ice sheet (Fig. 1d), with a similar cutoffs beyond 1500 m

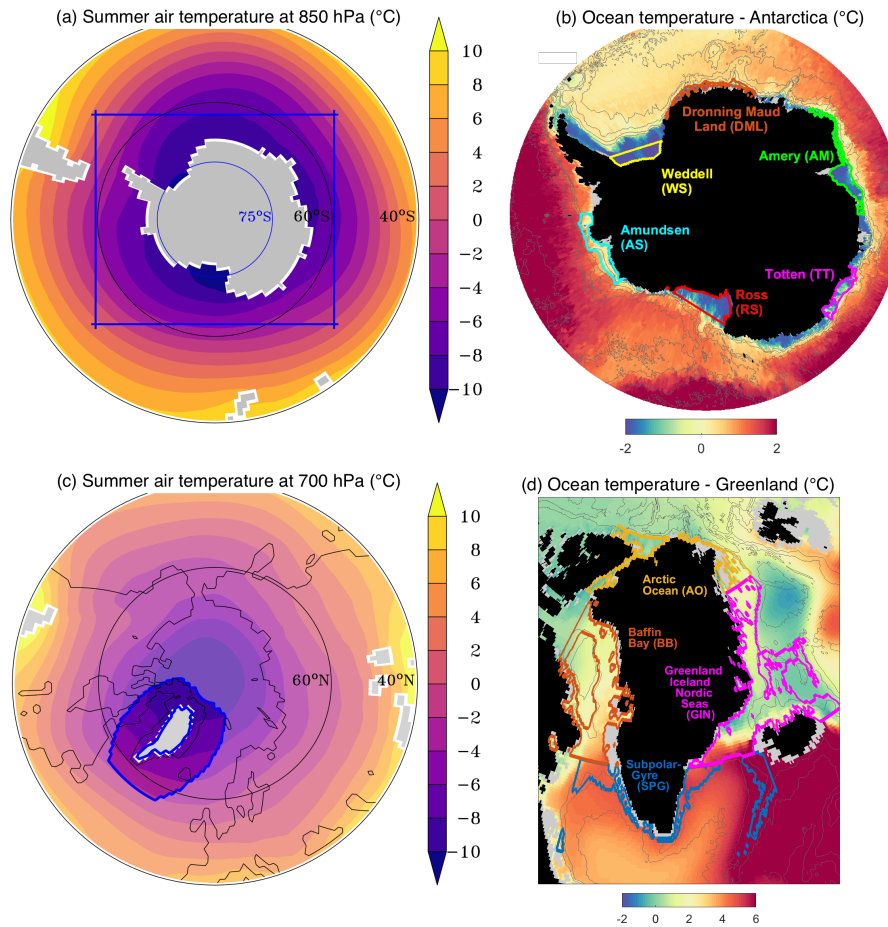


Figure 1. Atmosphere and ocean regions defined for metric computation. (a) For Antarctic atmosphere and surface ocean metrics, we considered the domain south of 40°S over ocean (color shading). The blue box shows standard lateral boundaries for regional climate models. Color shading is ERA-Interim summer air temperature at 850 hPa over 1980-2004. (b) For Antarctic ocean metrics, we considered 6 ocean sectors shallower than 1500 m. Color shading shows the depth-integrated temperature of our reference historical climatology. (c) For Greenland atmosphere metrics, we considered the domain inside the usual boundaries of MAR simulations in that region, i.e. inside the blue box, except where ice sheet topography is above 2000 m a.s.l. (bright color shading). Color shading is ERA-Interim summer air temperature at 700 hPa over 1980-2004. (d) For Greenland ocean metrics, we considered the 4 sectors shown with different colored outlines. Color shading shows the depth-integrated (200 to 500 m) temperature of our reference historical climatology.

bottom depth and geographical distance from the ice sheet to select coastal ocean cells near the ice sheet. For each sub-region, volume-averaged temperatures below 200 m depth are computed, providing a scalar near-shore sub-surface temperature metric. For Antarctica, the full depth range down to 1500 m is included, while for Greenland, the profiles are truncated below 500 m depth to account for shallow continental shelf depths and bottom sills that typically prevent inflows from greater depths toward the marine terminating glaciers in Greenland fjords ([Morlighem et al., 2017](#)).

Regional volume-averaged temperatures are also computed from available observed ocean climatologies, using the same
150 algorithm as for the model output. For Greenland, observational data are taken directly from the annually averaged statistical
fields of the 2013 World Ocean Atlas (~~Locarnini and Seidov, 2013~~)([WOA; Locarnini and Seidov, 2013](#)). For Antarctica, a re-
fined climatology of coastal water masses was constructed by combining the 2018 WOA data (~~WOA, Locarnini et al., 2019~~)
([Locarnini et al., 2019](#)) with statistical fields from the EN4 ocean climatology (Good et al., 2013) and publicly available tem-
perature profiles from Satellite Relay Data Logger–equipped seals (Roquet et al., 2018), with further details provided in
155 ~~Jourdain et al. (in prep)~~. [Nowicki \(2019\)](#). In both cases, ocean measurements close to the ice sheets are so sparse that all
observations are included in the computation of the regional averages, regardless of their acquisition date.

2.2.3 Aggregating historical metrics

In order to aggregate different metrics of varying nature and magnitude, each of the historical metrics ~~χ described above,~~
[described above \(denoted as \$\chi\$ below\)](#) is normalized with regards to the 33-model multi-model median and interquartile range
160 (IQR). For each model i :

$$\chi_{i, norm} = \frac{\chi_i - \text{median}(\chi)}{\text{IQR}(\chi)}. \quad (1)$$

We average the normalized metrics into three realms: atmosphere, surface ocean (for Antarctica), and sub-surface ocean.
This decision was made to weaken the dependence of the final ranking on the number of variables used for each realm.
Normalization of metrics prevents highly variable or large-amplitude metrics from being overly influential in the average (see
165 Fig. A.1) while still penalizing extremes. The final aggregated score for each model is obtained by averaging atmosphere and
ocean for Greenland, and atmosphere, surface ocean, and sub-surface ocean for Antarctica. An alternative aggregating method,
where all normalized metrics are weighted equally (12 for Antarctica, 7 for Greenland), is presented in Fig. A.2, and does not
change our conclusions.

2.3 Projected 21C changes

170 2.3.1 Climate change metrics

For atmospheric and surface ocean variables, climate change metrics are calculated as the difference between the 2070-2100
mean (RCP8.5) and the 1980-2010 mean (historical) value of each variable, spatially-averaged over the entire Greenland and
Antarctic atmospheric domains (Fig 1) ~~-, denoted with the Δ symbol. The only exception is for change in precipitable water,~~
[computed as the difference between 2070-2100 mean \(RCP8.5\) and the 1980-2010 mean \(historical\) divided by the 1980-2010](#)
175 [mean value of each variable, then spatially-averaged over the atmospheric domain, denoted with the \$\delta\$ symbol, because it](#)
[follows a log-normal distribution.](#) For the sub-surface ocean, we define metrics as the change in volume-averaged regional
temperature between the 1989-2009 and 2080-2100 periods. For Antarctica, we consider four metrics for the atmosphere
(change in annual air temperature at 850 hPa, $\Delta t_{a850}[a]$, in annual precipitable water, $\delta prw[a]$, and in position and strength
of the tropospheric westerly jet, ΔJ_{pos} and ΔJ_{str}), two metrics for the surface ocean (change in winter sea ice ~~concentration,~~

180 Δ se[fw]extent, Δ se[w]), and in summer sea surface temperature, Δ tos[s]), and six metrics for change in sub-surface ocean temperature (Δ T), one for each of the sectors defined in Section 2.2.2. For Greenland, we define two metrics for the atmosphere (change in annual air temperature at 700 hPa Δ ta700[a], and in annual precipitable water, δ prw[a]) and four metrics for change in sub-surface ocean temperature (Δ T), one for each ocean sector defined in Section 2.2.2.

2.3.2 Maximizing diversity of projected 21C changes

185 To maximize the diversity of future projections covered in a sub-selection of models of size n , we define the ensemble inter-model spread E by combining the pairwise model differences across the climate change metrics defined in Section 2.3.1 (12 for Antarctica, 6 for Greenland). The spread of a 3-model ensemble is computed as the following:

$$E_{n=3} = \sum_x |\chi_{model\ 1} - \chi_{model\ 2}| + |\chi_{model\ 2} - \chi_{model\ 3}| + |\chi_{model\ 1} - \chi_{model\ 3}|, \quad (2)$$

with χ the climate change metrics defined in Section 2.3.1. The ensemble that maximizes E for a given ensemble size n ($n = 3$ for top 3, $n = 6$ for top 6) is the one qualified as ‘most diverse’ in its future projections.

3 Results for Antarctica

In this section, we focus on the model selection for the Antarctic ice sheet, which is based on historical ranking (Section 3.1) and projection diversity (Section 3.2). The selected models are presented in Section 3.3.

3.1 Historical bias ranking

195 Over the Antarctic domain, the total normalized historical metric ranges between -0.32 (model of highest fidelity, CanESM2) and 1.50 (model of lowest fidelity, BMU-ESM), with a median value of 0.13 (Fig. 2a). Figure 2a shows the 33 climate models ranked by their historical metric, together with contributions of the sub-surface ocean, atmosphere (blue), atmosphere (orange) and surface ocean (yellow) to the total historical metric.

200 Models do not perform equally across the three realms. For example, GFDL-CM3 and EC-EARTH perform well in the atmosphere, with atmospheric metrics of -0.22 and -0.21 respectively, amongst the best models, but are ranked as low fidelity (with total bias scores of 0.46 and 0.54) due to their poor performance in ocean sub-surface and surface conditions. Conversely, IPSL-CM5B-LR performs well in the sub-surface ocean (metric of -0.20) but is penalized by its poor performance in the atmosphere (metric of 2.07) and surface ocean conditions (metric of 1.77).

205 Models also do not perform equally within each realm, indicating that biases originate due to regional processes for sub-surface ocean, or variable-specific biases for surface ocean and atmosphere. We provide the per-variable breakdown of the ocean sub-surface metric (Fig. 2b), and ocean surface and atmospheric metrics (Fig. 2c). Although this paper cannot address these differences in detail, we highlight a few notable sources of discrepancies between metrics. For example, the sub-surface heat in the Weddell Sea region is the largest single contributor to the ocean bias metric in several models (Fig. 2b), including EC-EARTH, MRI-CGM3 and BNU-ESM. The large ocean heat bias would warrant specific studies investigating the model

210 representation of the ocean climatology in that region. Similarly, in the atmosphere, precipitable water is the largest single bias for models such as IPSL-CM5B-LR, INM-CM4, and MRI-CGCM3, and would warrant further investigation to improve model representation of the historical period.

Models that perform better than the median (historical metric < 0.13) have reasonable values for all three realms: the worst metric for each realm is lower than 50% of the IQR away from the ensemble median for that realm (Fig. 2a). This result gives 215 confidence that these models have a good overall performance, rather than compensating biases across realms. Our averaging method was effective in penalizing models that have a low fidelity over an entire realm. For this reason, selecting the top 3 models in the top half of the 33 model ensures overall good performance of these models in both the ocean and atmosphere.

3.2 Projected changes

All 33 models considered in this study show an increase in air temperature over the Southern Ocean and Antarctic continent 220 between the end of the 21st century and the end of 20th century climatologies (Fig. 3a), with a ~~multimodel~~-multi-model mean increase of $2.54\text{ }^{\circ}\text{C}$. Nevertheless, the ensemble shows a spread of transient climate sensitivity, with an atmospheric warming ranging from $1.3\text{ }^{\circ}\text{C}$ (GFDL-ESM2G) to $3.6\text{ }^{\circ}\text{C}$ (BNU-ESM), with a median of $+2.5\text{ }^{\circ}\text{C}$. We highlight the 3 core (red) and 3 targeted (yellow) AOGCMs selected in Section 3.3, to illustrate the spread that they cover compared to the 33-model ensemble. Although the projected change in air temperature is only one of the variables we use to diagnose projected 225 atmospheric changes, it provides a good representation of projected changes in the atmosphere. Indeed, the changes in annual air temperature are strongly correlated ($R^2 > 0.82$) to the projected changes in seasonal air temperature, in annual and seasonal precipitable water, and strongly anti-correlated to changes in winter sea ice extent ($R^2 = 0.70$). Projected changes in wind jet strength, as quantified in Bracegirdle et al. (2018), show a weaker negative correlation with air temperature changes, although a decrease in jet strength is generally associated with a decrease in annual sea ice extent ($R^2 = 0.46$), as noted in Bracegirdle 230 et al. (2018).

Climate models also overwhelmingly project a 21st century increase in ocean temperatures around Antarctica. For example, the 33 models project a warming of the Amundsen shelf (Fig. 3b), ranging from no significant warming (lowest warming, MRI-CGCM3) to $+1.10\text{ }^{\circ}\text{C}$ (highest warming, IPSL-CM5B-LR), with a median value of $+0.45\text{ }^{\circ}\text{C}$. ~~In Fig. 3b, the models selected in the top 3 and top 6 are highlighted in red and yellow respectively, to illustrate the spread that they sample over the~~ 235 ~~Amundsen region.~~ Other regions show qualitatively similar range of projected changes, with the highest warming (as quantified by the median value of the ensemble) occurring in the Dronning Maud Land, Amery, and Totten regions (DML median: $+0.76\text{ }^{\circ}\text{C}$; Amery median: $+0.70\text{ }^{\circ}\text{C}$; Totten median: $+0.59\text{ }^{\circ}\text{C}$). The lowest projected warming occurs in the Weddell and Ross regions (Weddell median: $+0.21\text{ }^{\circ}\text{C}$; Ross median: $+0.30\text{ }^{\circ}\text{C}$). The Amundsen region, presented in Fig. 3b, is currently under scrutiny due to ice-shelf thinning and accelerating ice discharge in the last decade (Kimura et al., 2017; Mouginit et al., 2014; 240 Barletta et al., 2018), but ~~is only this region is~~ projected to warm moderately in the future according to the 33-model ensemble (Amundsen median: $+0.45\text{ }^{\circ}\text{C}$).

Unlike the atmospheric warming, which is a good proxy for other atmospheric changes, the projected ocean warming in the Amundsen region is only weakly correlated ($R^2 \leq 0.016$) to other ocean regions. Some significant correlation can be found

for neighboring regions in East Antarctica, such as between the Dronning Maud Land and Amery regions ($R^2 = 0.71$) and
245 between the Amery and Totten regions ($R^2 = 0.48$), but is low across other regions ($R^2 \leq 0.25$). Projected changes in the
ocean are relatively independent across regions (detailed in Fig. B.1), which confirms the added value of quantifying regional
ocean metrics rather than metrics integrated over all Antarctic shelves.

3.3 Recommended ensemble

3.3.1 Top 3 (core experiments)

250 In the case of the Antarctic domain, the selection criteria described in Section 2 led to 6 suitable coupled models ~~to choose from~~
(CanESM2, NorESM1-M, CSIRO-Mk3-6-0, CCSM4, MIROC-ESM-CHEM, MIROC-ESM), where availability of required
data from RCP2.6 projections is the strongest constraint. We then select the 3 models that maximize the ensemble diversity
 $E_{n=3}$, as defined in Section 2.3.2. The selection is robust to removing one of the metrics at a time and to changing the weight
of the metrics in the calculation (Appendix C1).

255 The top 3 models selected are, in alphabetical order, CCSM4 ([pink](#)), MIROC-ESM-CHEM ([red](#)), and NorESM1-M ([light
blue](#)). These 3 models sample different projected changes in Antarctica under the RCP8.5 scenario (Fig. 4a). Overall, NorESM1-
M shows a stronger end-of-21st-century ocean warming than the ensemble median ([dashed](#)), but a low atmospheric warming
compared to the model ensemble. Conversely, MIROC-ESM-CHEM features an ocean warming similar to that of the ensem-
ble median, associated with strong atmospheric changes, about one IQR higher than the median. Finally, CCSM4 shows very
260 distinct regional patterns of ocean warming, with strong warming in the Weddell and Totten regions, and lower warming in the
Ross and Dronning Maud Land regions, relative to the ensemble median. The projected atmospheric changes in CCSM4 are
on the high end of the ensemble, qualitatively similar to that of MIROC-ESM-CHEM. [The qualitative warming projected by
the 3 models selected for the Antarctic "core" experiments is summarized in Table 2.](#)

3.3.2 Top 6 (targeted experiments)

265 For the additional 3 models (targeted), CSIRO-Mk3-6-0 ([yellow](#)) is chosen because of its good ranking (Fig. 2) and median
projected changes (Fig. 3,4b), and is preferred to ACCESS1.0 ~~of similarly~~ ([which also shows](#) median projections under
RCP8.5) because of the availability of the RCP2.6 scenario. Each of the metrics of future change lies close to the multi-model
ensemble median (see Fig. 4b), meaning that approximately half of the 33 climate models predict higher changes than those of
CSIRO-Mk3-6-0, and half predict lower changes.

270 The other two models selected are, in alphabetical order, HadGEM2-ES ([brown](#)) and IPSL-CM5A-MR ([dark blue](#)). HadGEM2-
ES brings diversity to the 6-model ensemble because of its extreme end-of-21st-century warming in the ocean, particularly in
the Ross Sea. This extreme regional warming, more than 2 times larger than the IQR from the median value, is ruled out of
the top 3 because it is considered to be a less likely response than those produced by a [high large](#) number of distinct climate
models. Nevertheless, in an intercomparison effort such as ISMIP6, sampling high-end scenarios is essential to (i) examine
275 the response of ice-sheet models which may have run-away effects, (2) include high risk (low probability, high cost) scenarios

in terms of future sea level rise. The atmospheric changes produced by HadGEM2-ES are higher than the median, but not outliers. Finally, IPSL-CM5A-MR features an ocean warming lower than the ensemble median in most ocean regions, and atmospheric changes higher than the median. It is the only model selected with systematically low warming in the ocean, and can be thought of as the converse to NorESM1-M. Robustness of the model selection is demonstrated in Appendix C1. [The qualitative warming projected by the additional 3 models selected for the Antarctic "targeted" experiments is summarized in Table 2.](#)

4 Results for Greenland

In this section, we describe the model selection for the forcing of the Greenland ice sheet. The methods include the model evaluation (~~included below~~ [Section 4.1 and 4.2](#)) and ensemble selection ([Section 4.24.3](#)), mirroring the selection performed for the Antarctic ice sheet (Section 3).

4.1 Historical bias ranking

Coupled climate models do not perform equally over the sub-surface ocean and the atmosphere (Fig. 5a) around Greenland, consistent with finding for Antarctica, shown in Section 3. Some models perform well in the atmosphere but are penalized by their poor ocean performance. For example, CMCC-CMS is the median of the ensemble and features one of the lowest biases in the atmosphere (-0.69) and one of the highest biases in the ocean (0.73). Conversely, others perform well in the ocean but show high biases in the atmosphere (e.g. MRI-CGCM3). This unequal performance across the ocean and atmospheric variables supports the need to assess several components of coupled climate models together, rather than separately.

Investigating the source of biases in any given model is beyond the scope of this paper, which focuses on selecting 6 models suitable for the ISMIP6 simulations. Nevertheless, the ranking of the models can highlight significant biases. For example, the ocean bias in several models, most notably CMCC-CS, CMCC-CESM and IPSL-CM5B-LR, is dominated by a bias in ocean heat in the Arctic region. This large bias in temperature would warrant a specific study to improve model representation of that region. However, the observations in this region are scarce and we have a lower degree of confidence in the resulting ocean climatology in that region than in more frequently and densely observed regions, as discussed in Section 5.

The model ranking around Greenland highlights that the fidelity of coupled models is regionally dependent. The models of highest fidelity around Greenland do not necessarily perform well around Antarctica, and vice versa. For example, CanESM2 is the best-ranked model for Antarctica (see Section 3) but is ranked in the lower half of the ensemble around Greenland due in part to its ocean biases. Likewise, MIROC5 performs well on all metrics around Greenland, and has been extensively used in the relevant literature (e.g., Fettweis et al., 2013; Tedesco and Fettweis, 2012), but has strong atmospheric biases over Antarctica. Climate models are not expected to perform equally in all regions, nevertheless, it is important for the scientific community to keep those regional variations in mind, especially if using existing studies performed over a different region. This unequal performance across the Greenland and Antarctic regions also supports our decision to perform model ranking and selection independently for the two ice sheets.

Finally, the models that perform better than the median have ocean and atmosphere biases that lie lower than 0.5 IQR away from the median. Although biases in individual (regional) variables may be higher than that, this result confirms that the best ranked models have a good performance in both the sub-surface ocean and the atmosphere, and gives us confidence that the top half of the ensemble are suitable candidates for the Greenland model selection.

4.2 Future projection diversity

All 33 AOGCMs project atmospheric warming over Greenland by the end of the 21st century. Projections range from +1.95 °C (lowest warming, FIO-ESM) to +5.95 °C (highest warming, MIROC-ESM-CHEM) with a median warming of +4.09 °C (Fig. 6a). Models that made our final selection, highlighted in red (top 3) and yellow (top 6), sample a range of future warming. Similar to results presented for Antarctica (Section 3), the changes in annual air temperature over Greenland are a good proxy for most other atmospheric changes. Increase in 700hPa air temperature is associated with an ~~an~~ increase in precipitable water ($R^2 = 0.96$), an increase in ocean surface temperature ($R^2 = 0.60$), and a decrease in summer sea ice cover ($R^2 = 0.29$).

Most models also project an increase in ocean temperature on the shelf surrounding Greenland. Baffin Bay, for example, is projected to warm by +0.48 °C by the end of the 21st century, with models projecting between +0.07 °C (lowest warming, BCC-CSM1-1) and +1.70 °C (highest warming, CanESM2). The models selected in Section 4.3, highlighted in Fig. 6b, cover a range of projected warming in Baffin Bay. Two other regions show similar projected changes (Arctic median: +0.48°C; Subpolar Gyre (SPG) median: +0.49°C). The highest projected warming occurs in the Greenland-Iceland-Norwegian region (GIN), with a median warming of +0.76°C.

Projected changes across the ocean regions are correlated between the Arctic Ocean and GIN regions ($R^2 = 0.58$), and mildly correlated between the SPG and GIN regions ($R^2 = 0.31$). Other regions are only weakly correlated with each other (detailed in Fig. B.2), and ocean changes show no significant correlation with the projected atmospheric changes ($R^2 < 0.06$).

4.3 Recommended ensemble

In the case of Greenland, the availability of sub-daily ~~wind~~ outputs is a strong constraint for the model selection. This was a determining factor because existing studies over Greenland show that ~~the regional atmospheric model MAR outperforms RCMs outperform global~~ climate models in representing realistic surface mass balance (e.g., Noël et al., 2018; Fettweis et al., 2013).

4.3.1 Top 3 (core experiments)

When applying the selection criteria described in Section 2 and removing CNRM-CM5 and EC-EARTH due to unavailable data, 6 models remain for the top 3 selection (MIROC5, IPSL-CM5A-MR, NorESM1-M, ACCESS1-0, ACCESS1-3, HadGEM2-ES). In this case, MIROC5 was pre-selected, as it features changes similar to that of the ensemble median (dotted; Fig. 7a), meaning half of the models project stronger changes than those of MIROC5, and half project weaker changes. Two additional models are selected, maximizing ensemble diversity of three models ([MIROC5, model 1, model 2]).

The top 3 models selected are, in alphabetical order, HadGEM2-ES, MIROC5 and NorESM1-M. These 3 models show different patterns of projected changes by the end of the 21st century (Fig. 7a). As described above, MIROC5 is chosen as a good representation of the overall ensemble. HadGEM2-ES features high atmospheric changes, including increases in air temperature and precipitable water, of magnitude stronger than the ensemble median. The projected changes in ocean heat are more regionally dependent, with warming higher in the Arctic and GIN (north-east), and lower in Baffin Bay and SPG (south-west) relative to the ensemble median. Conversely, NorESM1-M features a warming in the atmosphere on the low-end of the 33-model ensemble projections. The ocean warming is also regionally dependent, with NorESM1-M featuring low warming in GIN, the Arctic, and the SPG regions and a strong warming in the Baffin Bay region. [The qualitative warming projected by the 3 models selected for the Greenland "core" experiments is summarized in Table 3.](#)

4.3.2 Top 6 ([models for the targeted experiments](#))

For the top 6 selection, 5 models (IPSL-CM5A-MR, CSIRO-Mk3-6-0, CCSM4, ACCESS1-0, ACCESS1-3) are available to complement the already selected top 3.

The selected models are, in alphabetical order, ACCESS1-3, CSIRO-Mk3-6-0, and IPSL-CM5A-MR. CSIRO-Mk3-6-0 projects a low atmospheric warming, far below the median value, alongside an extreme warming in the south-west ocean regions (ΔT Baffin-Bay-BB > 2 ; ΔT SPG = 0.94). ~~ACCESS1-3~~-ACCESS1-3 adds diversity to the ensemble as it shows strong warming in Baffin Bay and the Arctic Ocean, but low warming in the subpolar gyre region (SPG). Its atmospheric warming is close to the median. Finally, IPSL-CM5A-MR project strong warming in the Greenland-Iceland-Norwegian seas (GIN), while other ocean regions and atmospheric variables are closer to the median. [The qualitative warming projected by the additional 3 models selected as forcing for the Greenland "targeted" experiments is summarized in Table 3.](#)

5 Discussion

In this study, we evaluated the performance of 33 CMIP5 AOGCMs relative to reanalyses and gridded observational datasets covering the atmosphere, sea surface, and sub-surface ocean around the Greenland and Antarctic ice sheets. We also assessed 21st century changes in key oceanic and atmospheric variables. Time constraints for ISMIP6 simulations drove several decisions relating to the scope of this analysis, including: the use of the CMIP5 (rather than the now-partially-available CMIP6) ensemble; the use of AOGCMs that had already been processed and regridded for both the ocean and atmosphere; and the use of available observational products with limitations and biases, particularly in the ocean sub-surface. However, this assessment of near-ice sheet present-day and future climate remains the most comprehensive performed to date.

Many subjective choices were made in the model selection process. We have attempted to document these choices, and note that the relative insensitivity of results to alternate choices (e.g., Fig. ~~???~~-[Appendix A.2](#), [Appendix C](#)) provides some confidence that our rankings are robust for the CMIP5 ensemble. However, because the ~~rankings will not be applicable to evaluation and selection exercise will have to be repeated for~~ future model ensembles (e.g. CMIP6), our discussion focuses on key elements of our methodology that could be further developed. Implications are discussed with respect to results from the

full 33-member ensemble to extend the relevance to other exercises where the small ensemble required for ISMIP6 may not apply.

Model selection was made largely based on their representation of the present-day local climate, with the implicit assumption that biases relative to observations reflect a poor representation of processes of relevance to future warming. It is difficult to determine whether performance relative to this set of present-day regional metrics is: 1) a sufficient means to evaluate AOGCMs and 2) relevant to the rate of 21st century near-ice-sheet warming. Krinner and Flanner (2018) shows that model biases are stationary under future climate change within the CMIP5 dataset, providing justification for using less biased models for climate change studies. However, over the long timescales that ISMIP6 seeks to assess, different processes and/or biases (global and/or non-local ocean warming rates, e.g. stratospheric ozone recovery) may be equally important; i.e., even if a model closely matches historical conditions ~~it may be missing a key process important for projections~~, key processes for projections may still be missing.

Support for the relevance of these metrics might be derived from a clear relationship between the modern state and projections of change across models (so-called “emergent constraints”). Bracegirdle et al. (2015) and Agosta et al. (2015) found that 21st century changes in Antarctic air temperature and precipitation rate (and, perhaps surprisingly, jet strength (Bracegirdle et al., 2018)) were correlated to sea ice area bias across models. In this analysis, we found no significant correlation between historical biases and climate changes over Antarctica (or Greenland). A plausible explanation is our use of an 850 hPa (rather than surface) temperature metric and our circum-Antarctic study region. However, this result may also indicate a sensitivity to the specific models included in the ensemble: we find that the magnitude and significance of inter-model correlations are sensitive to whether all or a set of the best-performing models ~~is~~ are assessed. Shared code and parameterizations across models may also underlie some of the modest correlations ~~evident~~ evidenced in our analysis.

It is difficult to determine whether the historical metrics chosen in this analysis are comprehensive (e.g. account for all relevant processes) and/or independent. ~~With respect to~~ Concerning independence, we eliminated metrics which ~~represent~~ represent the same physical processes and are strongly correlated (e.g., the precipitation and air temperature variables in Bracegirdle et al. (2015) are strongly correlated to those in Agosta et al. (2015) and were not included in this study). Assessing comprehensiveness is more difficult. For example, the choice of metrics is constrained by the availability of observations. In particular, oceanographic measurements in the vicinity of ice sheets are very sparse and feature sharp horizontal gradients in water masses (e.g., Thompson et al., 2018). As a result, we chose to calculate volume- and time- mean quantities over subjectively defined regions in order to maximize the number of observations included. It is unclear which ocean region is most “important” in terms of future mass balance. The optimal number of regions, based on their relevance to future ice sheet change and their independence, remains to be determined. These choices should be expected to influence both evaluations of performance and warming. In contrast, observations for the atmosphere and surface ocean have better spatiotemporal coverage. Correspondingly, the metrics chosen were continental-scale and seasonally resolved. However, our continental-scale evaluation may obscure regional variability. Atmospheric dynamical modes, such as variability in the Amundsen Sea Low and the Southern Annular Mode (SAM), strongly impact the regional climate in Antarctica (Holland et al., 2017; Fyke et al., 2017). Although our grid-point error metric reflects biases in atmospheric pressure, it is not able to attribute the bias to a model’s

lack of fidelity to, say, the asymmetric nature of the SAM. Future work should more formally assess the number and relative weighting of regional metrics in the atmosphere and ocean, and include dynamically-relevant measures of asymmetry. Similar concerns apply to the metrics of future warming, and their relevance to ice sheet mass balance. We note that our analysis does not address the rate of warming, which differs widely across models. In the ocean, the rate and timing of warming may have dramatic effects on 21st century ice sheet evolution (Hellmer et al., 2012; Timmermann and Goeller, 2017).

We have noted the unequal performance of coupled climate models over different realms, which we suggest highlights the importance of assessing model fidelity over a range of metrics combining the sub-surface ocean, surface ocean and atmosphere conditions. It also explains why the present ranking of models differs from existing intercomparison studies specifically focused on the atmosphere (Agosta et al., 2015, e.g.,) (e.g., Agosta et al., 2015) or the ocean (e.g., Sallée et al., 2013; Meijers et al., 2012; Russell et al., 2018) (e.g., Meijers et al., 2012; Sallée et al., 2013; Russell et al., 2018). For example, the metrics used in Agosta et al. (2015) led to EC-EARTH and CanESM2 being ranked closely (8 and 9 out of 41 models), implying similar performance. However, by including the sub-surface ocean metrics, our results point to CanESM2 as the model with the best fidelity overall, while EC-EARTH is in the lower half of the 33-model ensemble due to its poor performance in the ocean (other examples of differences in rankings across realms can be found by examining Fig. 2 or Fig. 5). As Agosta et al. (2015) focuses purely on the model performance for ice-sheet surface mass balance, ~~its~~ their results differ from ~~this paper~~ the current study evaluating both the ocean and atmospheric metrics for the sake of providing the atmosphere-driven surface mass balance and the ocean-driven melt from the same coupled model as boundary conditions to ice-sheet models. This underscores the importance of considering the original aim of an intercomparison, including the variables and the regions considered, before interpreting or applying ranking derived from the analysis.

Antarctica and Greenland were treated independently, supported by the different model performance across the ensemble. A different set of models was selected for Greenland and Antarctica, suggesting model performance varies in polar regions of different hemispheres. However, with respect to future warming, it is reasonable to expect some degree of inter-hemispheric correlation in warming (e.g. due to a high AOGCM climate sensitivity). It is unclear how this inter-ice sheet independence assumption could influence sea level projections, as it depends upon the response of SMB and changes in ice flux ~~in~~ of the different ice sheets.

Using aggregated measures of present-day performance and future climate changes, we selected 6 AOGCMs as adequate and representative of future near-ice sheet warming pathways. This ensemble size was judged to be ~~feasible~~ reasonable for ISMIP6, given computational limitations and the goal to sample different sources of uncertainty (e.g. model, RCP scenarios, parameterizations, parameters values, etc). However, given the many degrees of freedom across the evaluation metrics, it is difficult to select a fully representative sample. Some limitations of the sample size are apparent, notably the non-uniform distribution across parameters (e.g. no low ocean warming sampled). Furthermore, the models selected are not structurally independent. For example, HadGEM2-ES and ~~ACCESS-1.3~~ ACCESS1-3 share a common Hadley Center atmospheric model, while NorESM1 and CCSM4 share the NCAR Community Atmospheric Model. Such interdependence may limit the diversity of forcing applied to ISMIP6 models. We do note that even if ISMIP6 had the ability to evaluate all available CMIP5 AOGCMS, issues with statistical sampling and diversity of CMIP models, code similarities/independence, and quality would persist (Knutti

et al., 2013; Sanderson et al., 2015a, b). Future model evaluation studies may invert the process used here: i.e., objectively assess the appropriate number of models to achieve sufficient diversity in forcing.

Finally, we emphasize that evaluation is only a first step to a better process-based understanding of the differences between models. It is critical to assess the processes that make models (or model families) perform better or ~~warm-at-a-different~~
445 ~~rate~~project climate warming at different rates. We invite modeling groups or researchers interested to examine these to trace back the source of the bias in individual models or across the larger ensemble.

6 Conclusions

As ~~a~~part of the Ice Sheet Model Intercomparison Project for CMIP6 (ISMIP6), ice sheet models will be forced with climate model-derived time series of basal melt (for Antarctica), front retreat (for Greenland) and surface mass balance. To generate
450 such forcing, a sub-set of CMIP5 models has been selected according to (i) their realistic representation of the historical period (as compared to reanalysis data), (ii) the diversity of the projected 21st-century changes under RCP8.5 within the selected subset. As a result of the evaluation and selection process ~~described in this paper~~performed in this study, six AOGCMS have been selected for ISMIP6 Antarctic future projection runs, ~~and~~including 3 for the "core" experiments (CCSM4; MIROC-ESM-CHEM; NorESM1-M) and 3 for the additional "targeted" experiments (CSIRO-Mk3-6-0; HadGEM2-ES; IPSL-CM5A-M) (see Table 2).
455 Independently, six AOGCMS have been selected for ISMIP6 Greenland future projection runs ~~-To complement the quantitative comparison described in the results section, a qualitative description of their projected warming is shown in Table 2, and (core experiments: HadGEM2-ES, MIROC5, NorESM1-M; targeted experiments: ACCESS1-3, CSIRO-Mk3-6-0, IPSL-CM5A-M; see Table 3-)~~. Ocean and atmospheric data from these AOGCMS is used to generate ice sheet surface mass balance, the Greenland retreat parameterization (e.g., Slater et al., 2019), and the Antarctic basal melt parameterization (Nowicki, 2019)
460 , which will be presented in detail in upcoming manuscripts. It is expected that the range of near-ice sheet climate changes simulated by these AOGCMS will result in diverse projections of ice sheet mass balance change when used to force ISMIP6 simulations. ~~Readers interested in the translation of AOGCM output to forcing for~~The evaluation and selection of models was a necessary first step to develop the current ISMIP6 simulations are encouraged to Slater et al. (2019), Jourdain et al. (in prep) and Nowicki et al. (in prep)experiment protocol, and can be improved upon for the next phase of ISMIP in multiple
465 ways. Firstly, future studies will evaluate and select models from the CMIP6 ensemble. Repeating this study with CMIP6 data will provide insight on whether new developments in climate models reduce ocean and atmospheric biases near ice-sheets. Secondly, results from the ice sheet simulations will provide insight on ice sheet model sensitivity. For example, future model selection may weight atmospheric changes more heavily than ocean changes if ice sheet models show a higher sensitivity to surface mass balance. In addition, future selection should look to include more dynamical metrics (e.g. Amundsen Sea
470 Low representation, ocean slope front position) and consider the rate of projected changes in the ensemble diversity. These improvements will ensure that, in an intercomparison project that remains computationally limited, we prioritize the forcing that is most fruitful.

The AOGCMs selected for ISMIP6 Greenland projection runs, and their qualitative projected warming, are summarized in Table 3.

475 **Appendix A: Robustness of historical ranking**

Appendix A provides additional information illustrating the robustness of the historical ranking methodology.

480 Firstly, we demonstrate the need to normalize the historical bias metrics (Fig. A.1). As the atmospheric bias metrics are based on variables as distinct as temperature, pressure, sea ice extent and precipitable water, the raw bias metrics have different mean values, and different inter-model variability (Fig. A.1a). By normalizing each metric by its ensemble median and interquartile range, the (normalized) metrics are scaled to cover a similar inter-model variability (Fig. A.1b) and can be combined into a single metric independently of the original magnitude of the raw variable.

485 Secondly, we illustrate the robustness of the historical ranking to the averaging method by providing an alternate ranking. In Figure A.2, the AOGCMs are ranked by averaging all (normalized) bias metrics with equal weight (dashed), instead of the bias metric used in this study (where each realm is given the same weight; black). For both the Antarctic (Fig. A.2a) and Greenland (Fig. A.2b) domains, the difference in ranking are minor, as only 2 models would switch between the top and bottom 50% of the 33-model ensemble, and neither of these models are present in the top 3 or top 6 ensembles.

Appendix B: Projected 21C ocean warming

490 Appendix B presents details of the projected ocean warming for each CMIP5 model between 1980-2000 and 2080-2100 under the RCP8.5 scenario. The warming over the 6 Antarctic shelf regions is presented in Figure B.1, while the warming over the 4 Greenland shelf regions is presented in Figure B.2. In each region, labels and markers (*, †) identify models selected in the top 3 and top 6 ensembles respectively. In each region, the majority of AOGCMs predicts a warming by the end of the 21st century, although the magnitude and inter-model spread of warming is regionally dependent.

Appendix C: Robustness of model selection

495 This appendix describes robustness of the model selection to modifications of the choice and weight of metrics. We repeat the model selection for the top 3 and top 6 models for Antarctica (Sec. 3.3) and Greenland (Sec. 4.3) under removal of one of the metrics at a time and under a change of the weighting. Overall, the model selection is robust to the described modifications.

C1 Robustness of Antarctic model selection top 3

Robustness of Antarctic model selection top 3

500 Table C.1 lists the selected model combinations with absolute and relative frequency of occurrence for the Antarctic top 3 model selection. The final model combination (NorESM1-M, MIROC-ESM-CHEM, CCSM4) occurs in 9 of 12 cases. One

additional model (CanESM2) is selected in 25 % of the cases. Table C.2 lists the absolute and relative occurrence of each individual model in the ~~combinations~~ possible combinations presented in Table C.1.

C1 Robustness of Antarctic model selection top 6

Robustness of Antarctic model selection top 6

505 Table C.3 lists the selected model combinations with absolute and relative frequency of occurrence for the Antarctic top 6 selection. The final model combination (NorESM1-M, MIROC-ESM-CHEM, CCSM4, CSIRO-Mk3-6-0, HadGEM2-ES, IPSL-CM5A-MR) occurs in 12 of 14 cases. Table C.4 lists the absolute and relative occurrence of each individual model in the combinations given in Table C.3. When equal weighting of the 14 metrics is applied, giving more emphasis on the surface ocean, HadGEM2-ES is still selected in 4 of 14 cases, but replaced by MPI-ESM-MR in the majority of cases (9 of 14).

510 C1 Robustness of Greenland model selection top 3

Robustness of Greenland model selection top 3

Table C.5 lists the selected model combinations with absolute and relative frequency of occurrence for the Greenland top 3 selection. The final model combination (MIROC5, NorESM1-M, HadGEM2-ES) was selected in all cases. Table C.6 lists the absolute and relative occurrence of each individual model in the combinations given in Table C.5.

515 The same results were obtained when metrics for the surface ocean ($\Delta\text{tos-sst}$ [a], $\Delta\text{sie-sie}$ [s], $\Delta\text{sie-sie}$ [w]) were added to the other metrics ($\Delta 700\text{hPa}$ [a], δprw [a], $\Delta\text{T SPG}$, $\Delta\text{T BB}$, $\Delta\text{T AO}$, $\Delta\text{T GIN}$).

C1 Robustness of Greenland model selection top 6

Robustness of Greenland model selection top 6

520 Table C.7 lists the selected model combinations with absolute and relative frequency of occurrence for the Greenland top 6 selection. The final model combination (MIROC5, IPSL-CM5A-MR, CSIRO-Mk3-6-0, NorESM1-M, HadGEM2-ES, ACCESS1-3) occurs in 7 of 9 cases, with CCSM4 replacing ACCESS1-3 in the remaining 2 cases. Table C.8 lists the absolute and relative occurrence of each individual model in the combinations given in Table C.7.

Similar results were obtained whether metrics for the surface ocean ($\Delta\text{tos-sst}$ [a], $\Delta\text{sie-sie}$ [s], $\Delta\text{sie-sie}$ [w]) were included or not.

525 *Author contributions.* AB, CA, CML, NJ, TH, HG, HS, FS and SN designed the study and the evaluation methodology. AB, CA and HG performed analysis on data provided by CA, CML, NJ, TH and TJB. AB prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Data availability. The supporting data is available at 10.5281/zenodo.3367347.

Acknowledgements. AB was supported by the U.S. Department of Energy (DOE) Office of Science Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling (EESM) program (HiLAT-RASM project), and the DOE Office of Science (Biological and Environmental Research), Early Career Research program. CA was supported by the Agence Nationale de la Recherche Scientifique, project ANR-15-CE01-0015 (AC-AHC2), and by the Fondation Albert 2 de Monaco under the project Antarctic-Snow (2018–2020). CML acknowledges financial support from NSF grants 1513396 and 1744792. TH acknowledges financial support from Norwegian Research Council projects 231549 and 280727. NJ's contribution was partly funded by the French National Research Agency (ANR) through the TROIS-AS (ANR-15-CE01-0005-01). HG has received funding from the program of the Netherlands Earth System Science Centre (NESSC), financially supported by the Dutch Ministry of Education, Culture and Science (OCW) under Grant nr. 024.002.001. HS was supported by grants from the NASA Cryospheric Science, Sea Level Change Team, and Modeling Analysis and Prediction Program. FS acknowledges financial support from NSF grants 1756272 and 1916566. TJB acknowledges support from NERC grant NE/N018486/1 and the AntClim21 Scientific Research Programme of the Scientific Committee on Antarctic Research. ~~All members~~
530 ~~of~~ We thank the Climate and Cryosphere (CliC) effort, which provided support for ISMIP6 through sponsoring of workshops, hosting the ISMIP6
540 ~~collaboration are thanked for website and wiki, and promoted ISMIP6. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP5. We thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the CMIP data and providing access, the University at Buffalo for ISMIP6 data distribution and upload, and the multiple funding agencies who
545 support CMIP5 and CMIP6 and ESGF. We thank the ISMIP6 steering committee, the ISMIP6 model selection group and ISMIP6 dataset preparation group for their continuous engagement in defining ISMIP6, and for discussions and feedback, with particular thanks to Donald Slater ~~,~~ and Denis Felikson.~~

References

- Agosta, C., Fettweis, X., and Datta, R.: Evaluation of the CMIP5 models in the aim of regional modelling of the Antarctic surface mass balance, *Cryosphere Discussions*, 9, 3113–3136, <https://doi.org/10.5194/tc-9-2311-2015>, 2015.
- Agosta, C., Amory, C., Kittel, C., Orsi, A., Favier, V., Gallée, H., Broeke, M. R. v. d., Lenaerts, J. T. M., Wessem, J. M. v., Berg, W. J. v. d., and Fettweis, X.: Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979–2015) and identification of dominant processes, *The Cryosphere*, 13, 281–296, <https://doi.org/10.5194/tc-13-281-2019>, 2019.
- Andrews, L. C., Catania, G. A., Hoffman, M. J., Gullely, J. D., Lüthi, M. P., Ryser, C., Hawley, R. L., and Neumann, T. A.: Direct observations of evolving subglacial drainage beneath the Greenland Ice Sheet, *Nature*, 514, 80, <https://doi.org/10.1038/nature13796>, 2014.
- Asay-Davis, X. S., Jourdain, N. C., and Nakayama, Y.: Developments in Simulating and Parameterizing Interactions Between the Southern Ocean and the Antarctic Ice Sheet, *Current Climate Change Reports*, 3, 316–329, <https://doi.org/10.1007/s40641-017-0071-0>, <https://doi.org/10.1007/s40641-017-0071-0>, 2017.
- Banwell, A. F., MacAyeal, D. R., and Sergienko, O. V.: Breakup of the Larsen B Ice Shelf triggered by chain reaction drainage of supraglacial lakes, *Geophysical Research Letters*, 40, 5872–5876, 2013.
- Barletta, V. R., Bevis, M., Smith, B. E., Wilson, T., Brown, A., Bordoni, A., Willis, M., Khan, S. A., Rovira-Navarro, M., Dalziel, I., Smalley, R., Kendrick, E., Konfal, S., Caccamise, D. J., Aster, R. C., Nyblade, A., and Wiens, D. A.: Observed rapid bedrock uplift in Amundsen Sea Embayment promotes ice-sheet stability, *Science*, 360, 1335–1339, <https://doi.org/10.1126/science.aao1447>, 2018.
- Bindschadler, R. A., Nowicki, S., Abe-Ouchi, A., Aschwanden, A., Choi, H., Fastook, J., Granzow, G., Greve, R., Gutowski, G., Herzfeld, U., Jackson, C., Johnson, J., Khroulev, C., Levermann, A., Lipscomb, W. H., Martin, M. A., Morlighem, M., Parizek, B. R., Pollard, D., Price, S. F., Ren, D., Saito, F., Sato, T., Seddik, H., Seroussi, H., Takahashi, K., Walker, R., and Wang, W. L.: Ice-sheet model sensitivities to environmental forcing and their use in projecting future sea level (the SeaRISE project), *Journal of Glaciology*, 59, 195–224, <https://doi.org/10.3189/2013JoG12J125>, 2013.
- Bracegirdle, T. J. and Marshall, G. J.: The Reliability of Antarctic Tropospheric Pressure and Temperature in the Latest Global Reanalyses, *Journal of Climate*, 25, 7138–7146, <https://doi.org/10.1175/JCLI-D-11-00685.1>, 2012.
- Bracegirdle, T. J., Stephenson, D. B., Turner, J., and Phillips, T.: The importance of sea ice area biases in 21st century multimodel projections of Antarctic temperature and precipitation, *Geophysical Research Letters*, 42, 10,832–10,839, <https://doi.org/10.1002/2015gl067055>, 2015.
- Bracegirdle, T. J., Hyder, P., and Holmes, C. R.: CMIP5 Diversity in Southern Westerly Jet Projections Related to Historical Sea Ice Area: Strong Link to Strengthening and Weak Link to Shift, *Journal of Climate*, 31, 195–211, <https://doi.org/10.1175/JCLI-D-17-0320.1>, 2018.
- Bromwich, D. H., Nicolas, J. P., and Monaghan, A. J.: An Assessment of Precipitation Changes over Antarctica and the Southern Ocean since 1989 in Contemporary Global Reanalyses *, *Journal of Climate*, 24, 4189–4209, <https://doi.org/10.1175/2011JCLI4074.1>, 2011.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- Dinniman, M. S., Asay-Davis, X. S., Galton-Fenzi, B. K., Holland, P. R., Jenkins, A., and Timmermann, R.: Modeling ice shelf/ocean interaction in Antarctica: A review, *Oceanography*, 29, 144–153, 2016.

- 585 Domack, E., Duran, D., Leventer, A., Ishman, S., Doane, S., McCallum, S., Amblas, D., Ring, J., Gilbert, R., and Prentice, M.: Stability of the Larsen B ice shelf on the Antarctic Peninsula during the Holocene epoch, *Nature*, 436, 681–685, <https://doi.org/10.1038/nature03908>, 2005.
- Dutheil, C., Bador, M., Lengaigne, M., Lefèvre, J., Jourdain, N. C., Vialard, J., Jullien, S., Peltier, A., and Menkes, C.: Impact of surface temperature biases on climate change projections of the South Pacific Convergence Zone, *Climate Dynamics*, pp. 1–23, <https://doi.org/10.1007/s00382-019-04692-6>, 2019.
- 590 Favier, V., Krinner, G., Amory, C., Gallée, H., Beaumet, J., and Agosta, C.: Antarctica-Regional Climate and Surface Mass Budget, *Current Climate Change Reports*, 3, 303–315, <https://doi.org/10.1007/s40641-017-0072-z>, <https://doi.org/10.1007/s40641-017-0072-z>, 2017.
- Ferry, N., Parent, L., Garric, G., Bricaud, C., Testut, C.-E., Galloudec, O. L., Lellouche, J.-M., Drévilion, M., Greiner, E., Barnier, B., Molines, J.-M., Jourdain, N., Guinehut, S., Cabanes, C., and Zawadzki, L.: GLORYS2V1 global ocean reanalysis of the altimetric era (1993–2009) at meso scale, *Mercator Ocean Quarterly Newsletter*, 44, 28–39, https://www.mercator-ocean.fr/wp-content/uploads/2015/05/Mercator-Ocean-newsletter-2012_44.pdf, 2012.
- 595 Fettweis, X., Franco, B., Tedesco, M., Van Angelen, J. H., Lenaerts, J. T., Van Den Broeke, M. R., and Gallée, H.: Estimating the Greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model MAR, *Cryosphere*, 7, 469–489, <https://doi.org/10.5194/tc-7-469-2013>, 2013.
- 600 Fettweis, X., Box, J. E., Agosta, C., Amory, C., Kittel, C., Lang, C., As, D. v., Machguth, H., and Gallée, H.: Reconstructions of the 1900–2015 Greenland ice sheet surface mass balance using the regional climate MAR model, *The Cryosphere*, 11, 1015–1033, <https://doi.org/10.5194/tc-11-1015-2017>, 2017.
- Franco, B., Fettweis, X., and Erpicum, M.: Future projections of the Greenland ice sheet energy balance driving the surface melt, *The Cryosphere*, 7, 1–18, <https://doi.org/10.5194/tc-7-1-2013>, 2013.
- 605 Frieler, K., Clark, P. U., He, F., Buizert, C., Reese, R., Ligtenberg, S. R. M., Broeke, M. R. v. d., Winkelmann, R., and Levermann, A.: Consistent evidence of increasing Antarctic accumulation with warming, *Nature Climate Change*, 5, 348–352, <https://doi.org/10.1038/nclimate2574>, 2015.
- Fyke, J., Lenaerts, J. T. M., and Wang, H.: Basin-scale heterogeneity in Antarctic precipitation and its impact on surface mass variability, *The Cryosphere (Online)*, 11, <https://doi.org/10.5194/tc-11-2595-2017>, 2017.
- 610 Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *Journal of Geophysical Research: Oceans*, 118, 6704–6716, <https://doi.org/10.1002/2013JC009067>, 2013.
- Gossart, A., Helsen, S., Lenaerts, J., Vanden Broucke, S., van Lipzig, N., and Souverijns, N.: An evaluation of surface climatology in state-of-the-art reanalyses over the Antarctic Ice Sheet, *Journal of Climate*, 2019.
- Hellmer, H. H., Kauker, F., Timmermann, R., Determann, J., and Rae, J.: Twenty-first-century warming of a large Antarctic ice-shelf cavity by a redirected coastal current, *Nature*, 485, <https://doi.org/10.1038/nature11064>, 2012.
- 615 Holland, M. M., Landrum, L., Kostov, Y., and Marshall, J.: Sensitivity of Antarctic sea ice to the Southern Annular Mode in coupled climate models, *Climate Dynamics*, 49, 1813–1831, <https://doi.org/10.1007/s00382-016-3424-9>, <https://doi.org/10.1007/s00382-016-3424-9>, 2017.
- Huai, B., Wang, Y., Ding, M., Zhang, J., and Dong, X.: An assessment of recent global atmospheric reanalyses for Antarctic near surface air temperature, *Atmospheric Research*, 226, 181–191, 2019.
- 620 Jourdain, N. e. a.: Ocean forcing for the ISMIP6 Antarctic ice sheet projections, *The Cryosphere*, in prep.

- Kendrick, A. K., Schroeder, D. M., Chu, W., Young, T. J., Christoffersen, P., Todd, J., Doyle, S. H., Box, J. E., Hubbard, A., Hubbard, B., Brennan, P. V., Nicholls, K. W., and Lok, L. B.: Surface Meltwater Impounded by Seasonal Englacial Storage in West Greenland, *Geophysical Research Letters*, 45, 10,474–10,481, <https://doi.org/10.1029/2018gl079787>, 2018.
- 625 Khan, S. A., Kjær, K. H., Bevis, M., Bamber, J. L., Wahr, J., Kjeldsen, K. K., Bjørk, A. A., Korsgaard, N. J., Stearns, L. A., van den Broeke, M. R., Liu, L., Larsen, N. K., and Muresan, I. S.: Sustained mass loss of the northeast Greenland ice sheet triggered by regional warming, *Nature Climate Change*, 4, 292–299, <https://doi.org/10.1038/nclimate2161>, 2014.
- Kimura, S., Jenkins, A., Regan, H., Holland, P. R., Assmann, K. M., Whitt, D. B., Van Wessem, M., van de Berg, W. J., Reijmer, C. H., and
630 Dutrieux, P.: Oceanographic Controls on the Variability of Ice-Shelf Basal Melting and Circulation of Glacial Meltwater in the Amundsen Sea Embayment, Antarctica, *Journal of Geophysical Research: Oceans*, 122, 10 131–10 155, <https://doi.org/10.1002/2017JC012926>, 2017.
- Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, *Geophysical Research Letters*, 40, 1194–1199, <https://doi.org/10.1002/grl.50256>, 2013.
- Krinner, G. and Flanner, M. G.: Striking stationarity of large-scale climate model bias patterns under strong climate change, *Proceedings of the National Academy of Sciences*, 115, 9462–9466, <https://doi.org/10.1073/pnas.1807912115>, 2018.
- 635 Krinner, G., Langeron, C., Ménégoz, M., Agosta, C., and Brutel-Vuilmet, C.: Oceanic Forcing of Antarctic Climate Change: A Study Using a Stretched-Grid Atmospheric General Circulation Model, *Journal of Climate*, 27, 5786–5800, <https://doi.org/10.1175/JCLI-D-13-00367.1>, 2014.
- Levermann, A., Winkelmann, R., Nowicki, S., Fastook, J. L., Frieler, K., Greve, R., Hellmer, H. H., Martin, M. A., Meinshausen, M., Mengel, M., Payne, A. J., Pollard, D., Sato, T., Timmermann, R., Wang, W. L., and Bindschadler, R. A.: Projecting Antarctic ice discharge using
640 response functions from SeaRISE ice-sheet models, *Earth System Dynamics*, 5, 271–293, <https://doi.org/10.5194/esd-5-271-2014>, 2014.
- Little, C. M. and Urban, N. M.: CMIP5 temperature biases and 21st century warming around the Antarctic coast, *Annals of Glaciology*, 57, 69–78, <https://doi.org/10.1017/aog.2016.25>, 2016.
- Locarnini, R. A., A. V. M. J. I. A. T. P. B. H. E. G. O. K. B. M. M. Z. C. R. P. J. R. R. D. R. J. M. H. and Seidov, D.: World Ocean Atlas 2013, Volume 1: Temperature, NOAA Atlas NESDIS 73, 2013.
- 645 Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., Reagan, J. R., Seidov, D., Weathers, K. W., Paver, C. R., and Smolyar, I.: World Ocean Atlas 2018, Volume 1: Temperature, NOAA Atlas NESDIS 81, https://data.nodc.noaa.gov/woa/WOA18/DOC/woa18_vol1.pdf, 2019.
- Meijers, A. J. S., Shuckburgh, E., Bruneau, N., Sallee, J.-B., Bracegirdle, T. J., and Wang, Z.: Representation of the Antarctic Circumpolar Current in the CMIP5 climate models and future changes under warming scenarios, *Journal of Geophysical Research: Oceans*, 117,
650 <https://doi.org/10.1029/2012JC008412>, 2012.
- Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., Catania, G., Chauché, N., Dowdeswell, J. A., Dorschel, B., et al.: BedMachine v3: Complete bed topography and ocean bathymetry mapping of Greenland from multibeam echo sounding combined with mass conservation, *Geophysical Research Letters*, 44, 11–051, 2017.
- Mouginot, J., Rignot, E., and Scheuchl, B.: Sustained increase in ice discharge from the Amundsen Sea Embayment, West Antarctica, from
655 1973 to 2013, *Geophysical Research Letters*, 41, 1576–1584, <https://doi.org/10.1002/2013GL059069>, 2014.
- Mouginot, J., Rignot, E., Bjørk, A. A., van den Broeke, M., Millan, R., Morlighem, M., Noël, B., Scheuchl, B., and Wood, M.: Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018, *Proceedings of the National Academy of Sciences*, 116, 9239–9244, <https://doi.org/10.1073/pnas.1904242116>, <https://www.pnas.org/content/116/19/9239>, 2019.

- Naughten, K. A., Meissner, K. J., Galton-Fenzi, B. K., England, M. H., Timmermann, R., Hellmer, H. H., Hattermann, T., and Debernard, J. B.: Intercomparison of Antarctic ice-shelf, ocean, and sea-ice interactions simulated by MetROMS-iceshelf and FESOM 1.4, *Geoscientific Model Development*, 11, 1257–1292, <https://doi.org/10.5194/gmd-11-1257-2018>, 2018.
- Noël, B., Fettweis, X., van de Berg, W. J., van den Broeke, M. R., and Erpicum, M.: Sensitivity of Greenland Ice Sheet surface mass balance to perturbations in sea surface temperature and sea ice cover: a study with the regional climate model MAR, *The Cryosphere*, 8, 1871–1883, <https://doi.org/10.5194/tc-8-1871-2014>, 2014.
- Noël, B., Van De Berg, W. J., Van Wessem, J. M., Van Meijgaard, E., Van As, D., Lenaerts, J. T., Lhermitte, S., Munneke, P. K., Smeets, C. J., Van Ulf, L. H., Van De Wal, R. S., and Van Den Broeke, M. R.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2 - Part 1: Greenland (1958-2016), *Cryosphere*, 12, 811–831, <https://doi.org/10.5194/tc-12-811-2018>, 2018.
- Nowicki, S.: ISMIP6 wiki page, http://www.climate-cryosphere.org/wiki/index.php?title=ISMIP6_wiki_page#ISMIP6_Standalone_Ice_Sheet_Experiments, 2019.
- Nowicki, S. M. J., Payne, A., Larour, E., Seroussi, H., Goelzer, H., Lipscomb, W., Gregory, J., Abe-Ouchi, A., and Shepherd, A.: Ice Sheet Model Intercomparison Project (ISMIP6) contribution to CMIP6, *Geoscientific Model Development*, 9, 4521–4545, <https://doi.org/10.5194/gmd-9-4521-2016>, 2016.
- Palermé, C., Genthon, C., Claud, C., Kay, J. E., Wood, N. B., and L’Ecuyer, T.: Evaluation of current and projected Antarctic precipitation in CMIP5 models, *Climate dynamics*, 48, 225–239, 2017.
- Pattyn, F., Perichon, L., Durand, G., Favier, L., Gagliardini, O., Hindmarsh, R. C., Zwinger, T., Albrecht, T., Cornford, S., Docquier, D., Fürst, J. J., Goldberg, D., Gudmundsson, G. H., Humbert, A., Hütten, M., Huybrechts, P., Jouvét, G., Kleiner, T., Larour, E., Martin, D., Morlighem, M., Payne, A. J., Pollard, D., Rückamp, M., Rybak, O., Seroussi, H., Thoma, M., and Wilkens, N.: Grounding-line migration in plan-view marine ice-sheet models: Results of the ice2sea MISMIP3d intercomparison, *Journal of Glaciology*, 59, 410–422, <https://doi.org/10.3189/2013JoG12J129>, 2013.
- Rignot, E., Casassa, G., Gogineni, P., Krabill, W., Rivera, A., and Thomas, R.: Accelerated ice discharge from the Antarctic Peninsula following the collapse of Larsen B ice shelf, *Geophysical Research Letters*, 31, <https://doi.org/10.1029/2004gl020697>, 2004.
- Roquet, F., Guinet, C., Charrassin, J.-B., Costa, D. P., Kovacs, K. M., Lydersen, C., Bornemann, H., Bester, M. N., Muelbert, M. C., Hindell, M. A., McMahon, C. R., Harcourt, R., Boehme, L., and Fedak, M. A.: MEOP-CTD in-situ data collection: a Southern ocean Marine-mammals calibrated sea water temperatures and salinities observations, <https://doi.org/10.17882/45461>, 2018.
- Russell, J. L., Kamenkovich, I., Bitz, C., Ferrari, R., Gille, S. T., Goodman, P. J., Hallberg, R., Johnson, K., Khazmutdinova, K., Marinov, I., Mazloff, M., Riser, S., Sarmiento, J. L., Speer, K., Talley, L. D., and Wanninkhof, R.: Metrics for the evaluation of the southern ocean in coupled climate models and earth system models, *Journal of Geophysical Research: Oceans*, 123, 3120–3143., <https://doi.org/10.1002/2017JC013461>, 2018.
- Sallée, J., Shuckburgh, E., Bruneau, N., Meijers, a. J. S., Bracegirdle, T. J., Wang, Z., and Roy, T.: Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: Historical bias and forcing response, *Journal of Geophysical Research: Oceans*, 118, 1830–1844, <https://doi.org/10.1002/jgrc.20135>, 2013.
- Sanderson, B. M., Knutti, R., and Caldwell, P.: Addressing Interdependency in a Multimodel Ensemble by Interpolation of Model Properties, *Journal of Climate*, 28, 5150–5170, <https://doi.org/10.1175/JCLI-D-14-00361.1>, 2015a.
- Sanderson, B. M., Knutti, R., and Caldwell, P.: A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble, *Journal of Climate*, 28, 5171–5194, <https://doi.org/10.1175/JCLI-D-14-00362.1>, 2015b.

- Schmidtko, S., Heywood, K. J., Thompson, A. F., and Aoki, S.: Multidecadal warming of Antarctic waters, *Science*, 346, 1227–1231, <https://doi.org/10.1126/science.1256117>, 2014.
- 700 Shepherd, A., Ivins, E., Rignot, E., Smith, B., Broeke, M. v. d., Velicogna, I., Whitehouse, P., Briggs, K., Joughin, I., Krinner, G., Nowicki, S., Payne, T., Scambos, T., Schlegel, N., A. G., Agosta, C., Ahlstrøm, A., Babonis, G., Barletta, V., Blazquez, A., Bonin, J., Csatho, B., Cullather, R., Felikson, D., Fettweis, X., Forsberg, R., Gallee, H., Gardner, A., Gilbert, L., Groh, A., Gunter, B., Hanna, E., Harig, C., Helm, V., Horvath, A., Horwath, M., Khan, S., Kjeldsen, K. K., Konrad, H., Langen, P., Lecavalier, B., Loomis, B., Luthcke, S., McMillan, M., Melini, D., Mernild, S., Mohajerani, Y., Moore, P., Mouginot, J., Moyano, G., Muir, A., Nagler, T., Nield, G., Nilsson, J., Noel, B., Otosaka, I., Pattle, M. E., Peltier, W. R., Pie, N., Rietbroek, R., Rott, H., Sandberg-Sørensen, L., Sasgen, I., Save, H., Scheuchl, B., Schrama, E., Schröder, L., Seo, K.-W., Simonsen, S., Slater, T., Spada, G., Sutterley, T., Talpe, M., Tarasov, L., Berg, W. J. v. d., Wal, W. v. d., Wesse, M. v., Vishwakarma, B. D., Wiese, D., and Wouters, B.: Mass balance of the Antarctic Ice Sheet from 1992 to 2017, *Nature*, 558, 219–222, <https://doi.org/10.1038/s41586-018-0179-y>, 2018.
- 705 Slater, D., Straneo, F., Felikson, D., Little, C., Goelzer, H., Fettweis, X., and Holte, J.: Past and future response of Greenland’s tidewater glaciers to submarine melting, *The Cryosphere Discussions*, 2019, 1–32, <https://doi.org/10.5194/tc-2019-98>, 2019.
- Stewart, A. L., Klocker, A., and Menemenlis, D.: Circum-Antarctic shoreward heat transport derived from an eddy-and tide-resolving simulation, *Geophysical Research Letters*, 45, 834–845, 2018.
- 710 Straneo, F. and Cenedese, C.: The Dynamics of Greenland’s Glacial Fjords and Their Role in Climate, *Annual Review of Marine Science*, 7, 89–112, <https://doi.org/10.1146/annurev-marine-010213-135133>, PMID: 25149564, 2015.
- Straneo, F., Sutherland, D. A., Holland, D., Gladish, C., Hamilton, G. S., Johnson, H. L., Rignot, E., Xu, Y., and Koppes, M.: Characteristics of ocean waters reaching Greenland’s glaciers, *Annals of Glaciology*, 53, 202–210, <https://doi.org/10.3189/2012AoG60A059>, 2012.
- 715 Tedesco, M. and Fettweis, X.: 21st century projections of surface mass balance changes for major drainage systems of the Greenland ice sheet, *Environmental Research Letters*, 7, <https://doi.org/10.1088/1748-9326/7/4/045405>, 2012.
- Thompson, A. F., Stewart, A. L., Spence, P., and Heywood, K. J.: The Antarctic Slope Current in a Changing Climate, *Reviews of Geophysics*, 56, 741–770, <https://doi.org/10.1029/2018RG000624>, 2018.
- 720 Timmermann, R. and Goeller, S.: Response to Filchner–Ronne Ice Shelf cavity warming in a coupled ocean–ice sheet model–Part 1: The ocean perspective, *Ocean Science*, 13, 765–776, 2017.
- Trusel, L. D., Frey, K. E., Das, S. B., Munneke, P. K., and Van Den Broeke, M. R.: Satellite-based estimates of Antarctic surface meltwater fluxes, *Geophysical Research Letters*, 40, 6148–6153, 2013.
- Trusel, L. D., Frey, K. E., Das, S. B., Karnauskas, K. B., Munneke, P. K., Van Meijgaard, E., and Van Den Broeke, M. R.: Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios, *Nature Geoscience*, 8, 927, 2015.
- 725 Trusel, L. D., Das, S. B., Osman, M. B., Evans, M. J., Smith, B. E., Fettweis, X., McConnell, J. R., Noël, B. P. Y., and Broeke, M. R. v. d.: Nonlinear rise in Greenland runoff in response to post-industrial Arctic warming, *Nature*, 564, 104–108, <https://doi.org/10.1038/s41586-018-0752-4>, 2018.
- van den Broeke, M.: Strong surface melting preceded collapse of Antarctic Peninsula ice shelf, *Geophysical Research Letters*, 32, 2005.
- 730 van Wesse, J. M., van de Berg, W. J., Noël, B. P. Y., van Meijgaard, E., Amory, C., Birnbaum, G., Jakobs, C. L., Krüger, K., Lenaerts, J. T. M., Lhermitte, S., Ligtenberg, S. R. M., Medley, B., Reijmer, C. H., van Tricht, K., Trusel, L. D., van Uft, L. H., Wouters, B., Wuite, J., and van den Broeke, M. R.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2 – Part 2: Antarctica (1979–2016), *The Cryosphere*, 12, 1479–1498, <https://doi.org/10.5194/tc-12-1479-2018>, 2018.

- Velicogna, I.: Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE, *Geophysical Research Letters*, 36, 2–5, <https://doi.org/10.1029/2009GL040222>, 2009.
- 735 Yin, J., Overpeck, J. T., Griffies, S. M., Hu, A., Russell, J. L., and Stouffer, R. J.: Different magnitudes of projected subsurface ocean warming around Greenland and Antarctica, *Nature Geoscience*, 4, 524, <https://doi.org/10.1038/ngeo1189>, 2011.
- Zwally, H. J., Li, J., Brenner, A. C., Beckley, M., Cornejo, H. G., DiMarzio, J., Giovinetto, M. B., Neumann, T. A., Robbins, J., Saba, J. L., Yi, D., and Wang, W.: Greenland ice sheet mass balance: distribution of increased mass loss with climate warming; 2003–07 versus 1992–2002, *Journal of Glaciology*, 57, 88–102, <https://doi.org/10.3189/002214311795306682>, 2011.

Table 1. ERA-Interim reanalysis and CMIP5 models used in this study.

Name	Modeling group	Atm. grid spacing	6-hourly available	rcp26 available	rcp85 available
ERA-Interim	ECMWF	0.7°	x		
ACCESS1-0	CSIRO-BOM	1.25°	x		x
ACCESS1-3	CSIRO-BOM	1.25°	x		x
BCC-CSM1-1	BCC	2.8°		x	x
BNU-ESM	GCESS	2.8°		x	x
CanESM2	CCCma	2.8°	x	x	x
CCSM4	NSF-DOE-NCAR	1.25°	x	x	x
CESM1-BGC	NSF-DOE-NCAR	1.25°			x
CESM1-CAM5	NSF-DOE-NCAR	1.25°		x	x
CMCC-CESM	CMCC	3.75°		x	x
CMCC-CM	CMCC	0.75°			x
CMCC-CMS	CMCC	1.8°			x
CNRM-CM5	CNRM-CERFACS	1.4°	x	x	x
CSIRO-Mk3-6-0	CSIRO-QCCCE	1.9°	x	x	x
EC-EARTH	EC-EARTH	1.125°		x	x
FGOALS-g2	LASG-IAP	2.8°		x	x
FIO-ESM	FIO	2.875°		x	x
GFDL-CM3	NOAA GFDL	1.8°	x	x	x
GFDL-ESM2G	NOAA GFDL	2.0°	x	x	x
GFDL-ESM2M	NOAA GFDL	2.0°	x	x	x
HadGEM2-CC	MOHC	1.25°			
HadGEM2-ES	MOHC	1.25°	x	x	x
INM-CM4	INM	1.5°			x
IPSL-CM5A-LR	IPSL	1.9°	x	x	x
IPSL-CM5A-MR	IPSL	1.3°	x	x	x
IPSL-CM5B-LR	IPSL	1.3°	x		x
MIROC-ESM	MIROC	2.8°	x	x	x
MIROC-ESM-CHEM	MIROC	2.8°	x	x	x
MIROC5	MIROC	1.4°	x	x	x
MPI-ESM-LR	MPI-M	1.9°		x	x
MPI-ESM-MR	MPI-M	1.8°		x	x
MRI-CGCM3	MRI	1.1°	x	x	x
NorESM1-M	NCC	1.9°	x	x	x
NorESM1-ME	NCC	1.9°		x	x

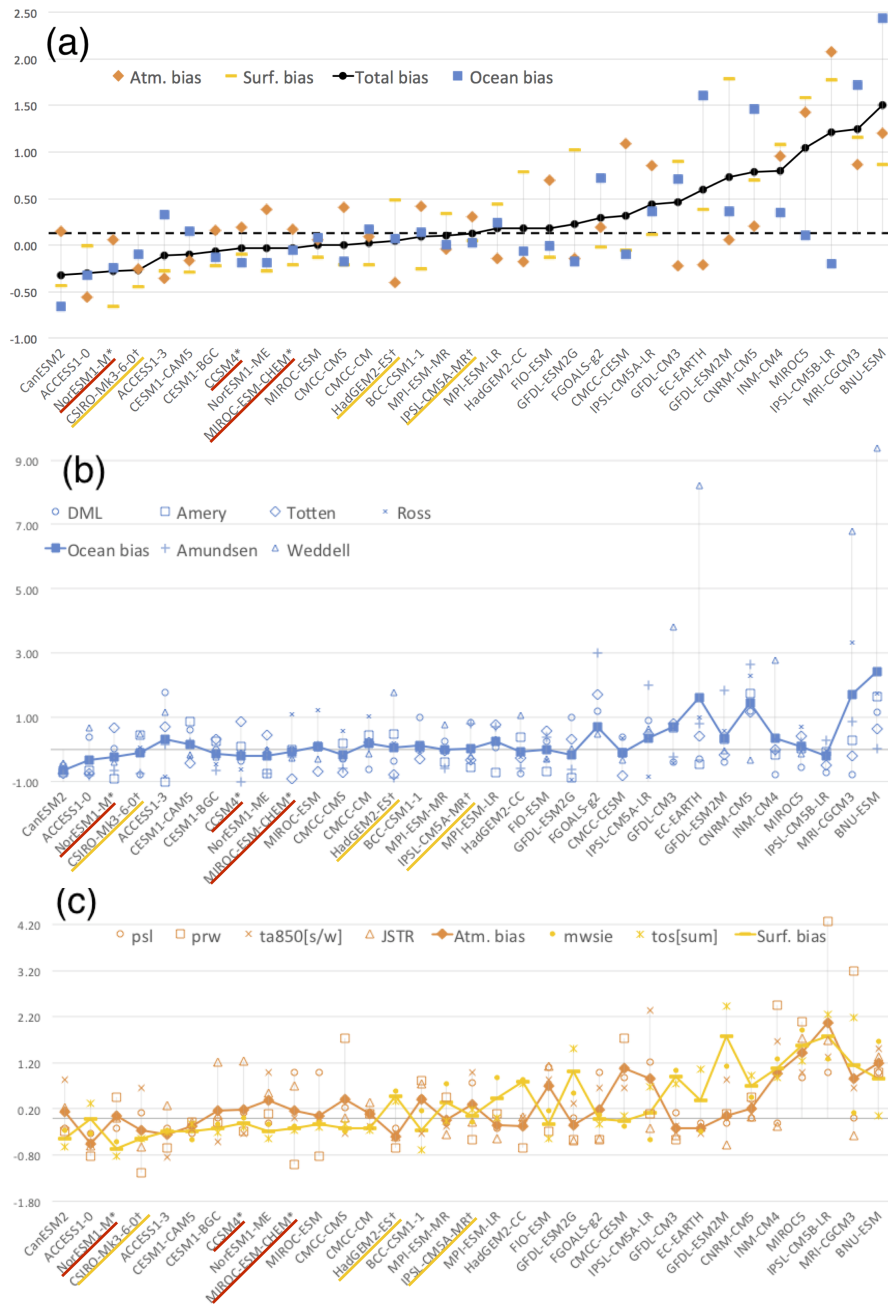


Figure 2. (a) Ranking of models according to total bias (black) over the Antarctic domain, with a break-down of the ocean (blue), atmosphere (orange) and surface (yellow) contributions. (b) Break-down of model performance in the ocean over the Antarctic domain. (c) Break-down of model performance in the atmosphere (orange) and ocean surface (yellow) over the Antarctic domain. Models are ranked according to total bias. Models selected in the top 3 (core) and top 6 (targeted) ensembles are underlined in color (red, yellow) and with markers (*, †) respectively.

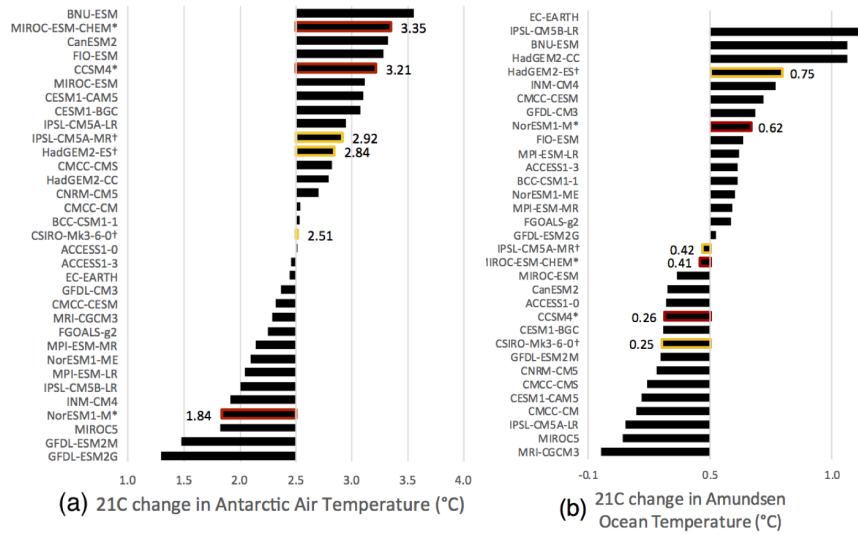


Figure 3. Projected RCP8.5 warming for each CMIP5 model in the Antarctic region. (a) Change in 850 hPa air temperature over the Southern ocean between 1980-2000 and 2080-2100. (b) Change in ocean temperature in the Amundsen region between 1980-2000 and 2080-2100. Models selected in the top3 (top6) ensemble are highlighted in red (yellow).

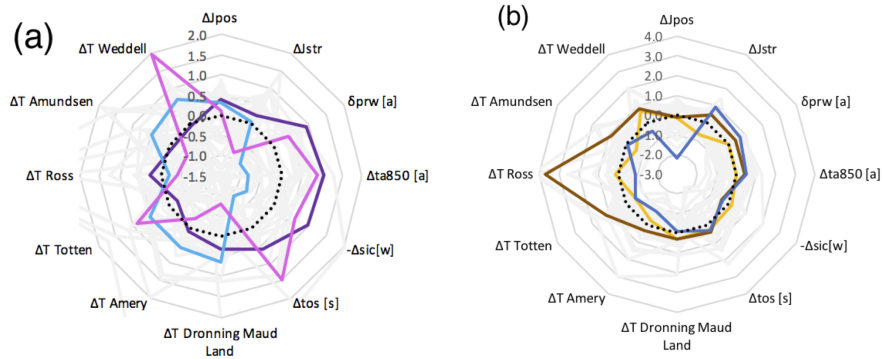


Figure 4. Normalized projected 21C changes for Antarctica (with model ensemble in gray and the median in black). (a) Top 3: CCSM4 (pink), MIROC-ESM-CHEM (purple), and NorESM1-M (light blue); (b) Top 4-6: CSIRO-Mk3-6-0 (yellow), HadGEM2-ES (brown) and IPSL-CM5A-M (dark blue).

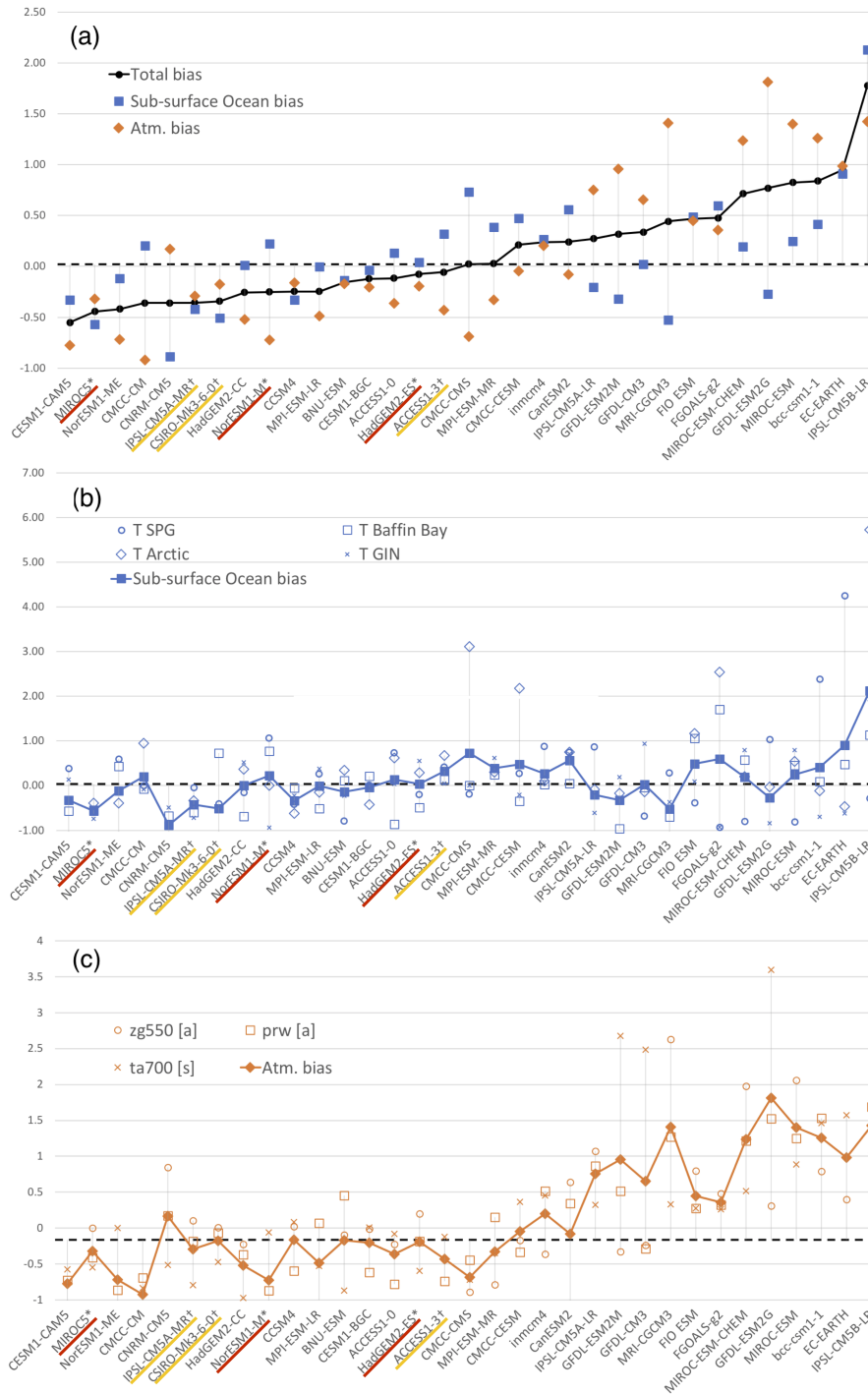


Figure 5. (a) Ranking of models according to total bias (black) over the Greenland domain, including the ocean (blue) and atmosphere (orange) contributions. (b) Break-down of model performance in the ocean over the Greenland domain. (c) Break-down of model performance in the atmosphere over the Greenland domain. Models are ranked according to total bias. Models selected in the top 3 (core) and top 6 (targeted) ensembles are underlined in color (red, yellow) and with markers (*, †) identify models selected in the top 3 and top 6 ensembles respectively.

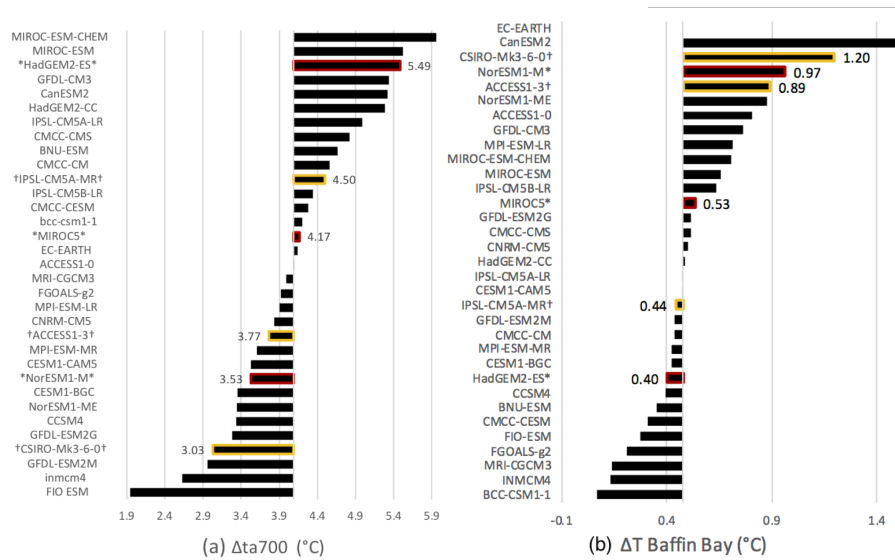


Figure 6. Projected RCP8.5 warming for each CMIP5 model over Greenland. (a) Change in 700 hPa air temperature over the Southern ocean between 1980-2000 and 2080-2100. (b) Change in ocean temperature in the Baffin Bay region between 1980-2000 and 2080-2100. Models selected in the top3 (top6) ensemble are highlighted in red (yellow).

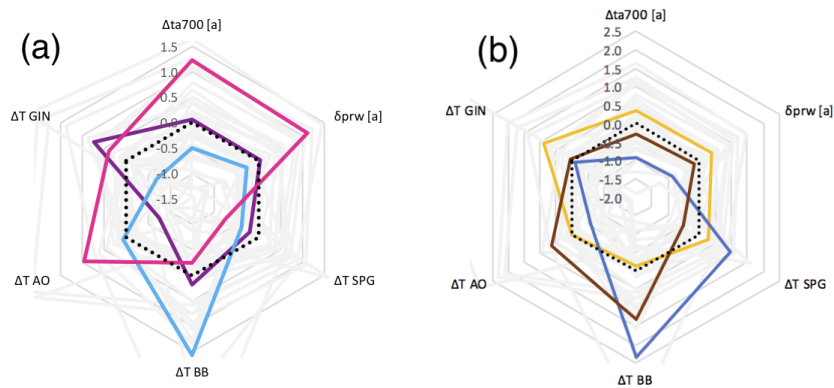


Figure 7. Normalized projected 21C changes for Greenland (with model ensemble in gray and the median in black). (a) Top 3: HadGEM2-ES (pink), MIROC5 (purple) and NorESM1-M (light blue); (b) Top 4-6: ACCESS1-3 (brown), CSIRO-Mk3-6-0 (dark blue), and IPSL-CM5A-MR (yellow). Ocean warming is calculated over 4 sectors (BB = Baffin Bay; AO = Arctic Ocean; GIN = Greeland-Iceland-Norwegian Seas; SPG = Subpolar Gyre).

Table 2. Selected AOGCMs for Antarctica and their qualitative projected warming

Model	Ocean	Atmosphere	Comments
CCSM4	median	high	Strong regional ocean differences
MIROC-ESM-CHEM	median	high	
NorESM1-M	mid-to-high	low	
CSIRO-Mk3-6-0	median	median	Extreme warming in the Ross Sea
HadGEM2-ES	high	median	
IPSL-CM5A-M	low	high	

Table 3. Selected AOGCMs for Greenland and their qualitative projected warming

Model	Ocean	Atmosphere	Comments
HadGEM2-ES	high	low	Strong warming in Baffin Bay
MIROC5	median	median	
NorESM1-M	median	high	Strong warming in the Arctic Ocean
ACCESS1-3 <u>ACCESS1-3</u>	mid-to-high	median	Strong warming in Baffin Bay
CSIRO-Mk3-6-0	mid-to-high	low	Extreme warming in Baffin Bay
IPSL-CM5A-M	low	high	Strong warming in the GIN Seas

Table C.1. ~~Top 3 selected~~ Possible model combinations for the Antarctica top 3 selection, with absolute and relative frequency of occurrence ~~in~~ when applying the robustness test

Model combination	Count	Occurrence
NorESM1-M, MIROC-ESM-CHEM, CCSM4	9	0.75
CanESM2, NorESM1-M, CCSM4	3	0.25

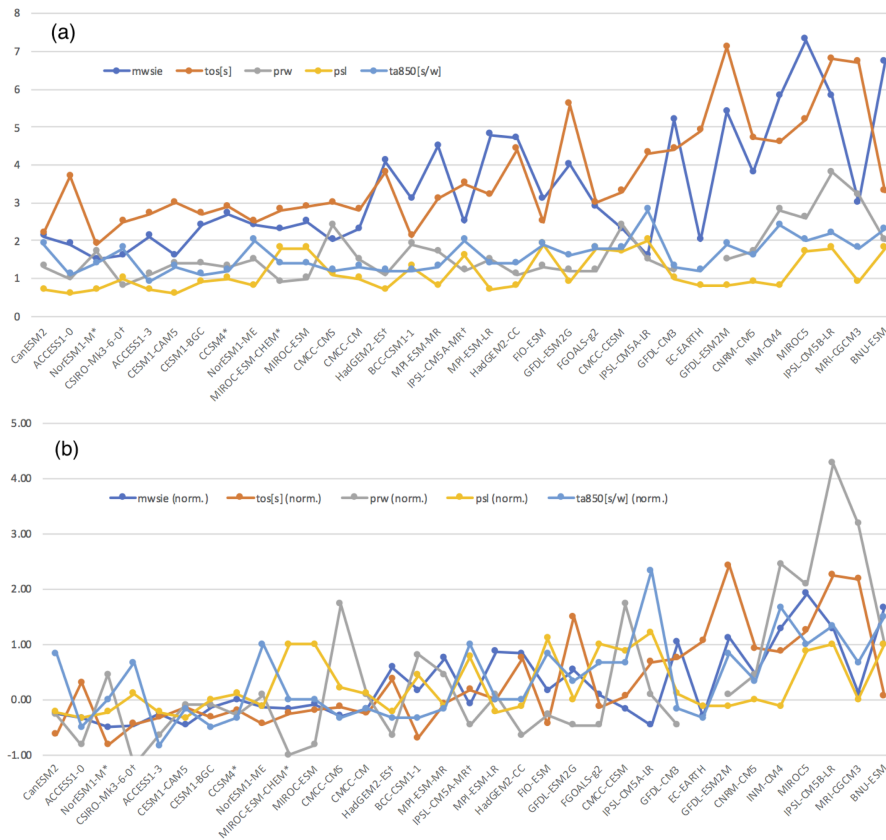


Figure A.1. Normalization of variables: (a) Historical atmospheric biases for the Antarctic domain; (b) Normalized historical atmospheric biases for the Antarctic domain. The non-normalized variables have different mean values, and different variability. The normalization removes the offset and rescales the variability, so that variables of different nature, magnitude, and variability can be combined in one atmospheric bias metric. Labels and markers (*, †) identify models selected in the top3 and top6 ensembles respectively.

Table C.2. Top 3 selected models for Antarctica with absolute Absolute and relative occurrence of each individual model included in the Antarctica top 3 combinations (Tab-presented in Table C.1)

Models	Count	Occurrence
NorESM1-M	12	1.00
CCSM4	12	1.00
MIROC-ESM-CHEM	9	0.75
CanESM2	3	0.25

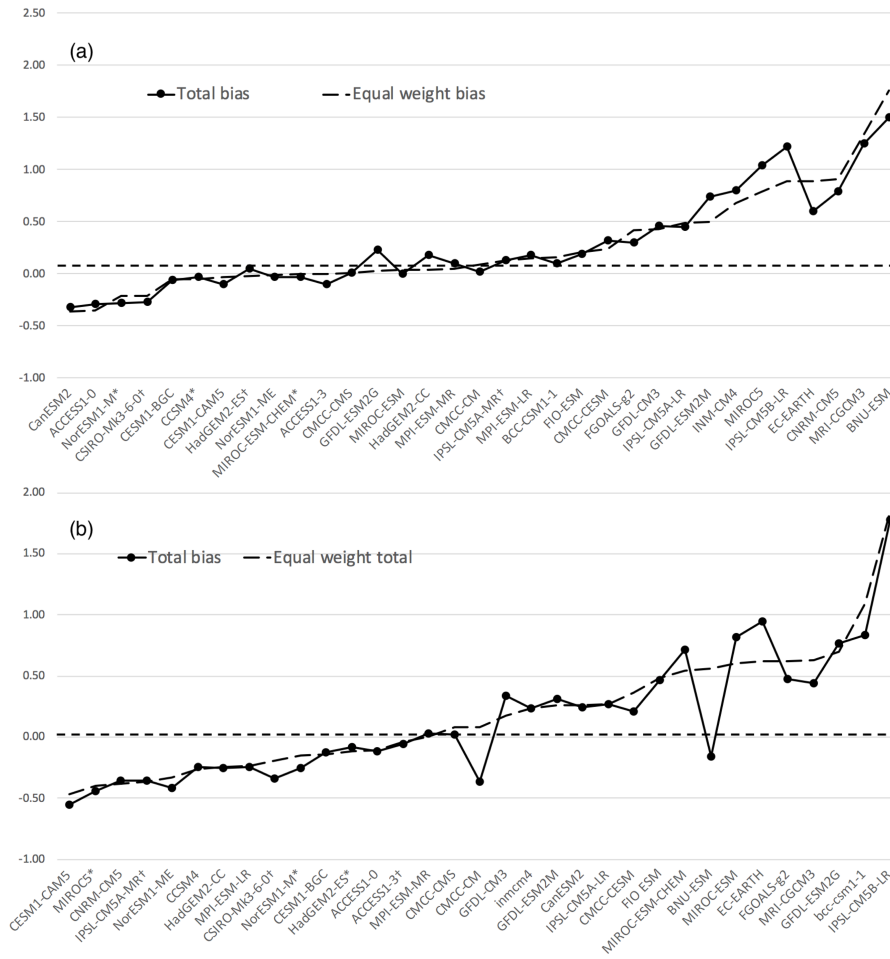


Figure A.2. Alternate ranking of AOGCMs according to an equal-weight total bias (dashed black) compared to the realms-averaged total bias (black) over (a) the Antarctic domain, (b) the Greenland domain. Labels and markers (*, †) identify models selected in the top3 and top6 ensembles respectively.

Table C.3. Top 6 selected Possible model combinations for the Antarctica top 6 selection, with absolute and relative frequency of occurrence in when applying the robustness test.

Model combination	Count	Occurrence
NorESM1-M, MIROC-ESM-CHEM, CCSM4, CSIRO-Mk3-6-0, HadGEM2-ES, IPSL-CM5A-MR	12	0.86
NorESM1-M, MIROC-ESM-CHEM, CCSM4, CSIRO-Mk3-6-0, BCC-CSM1-1, IPSL-CM5A-MR	1	0.07
NorESM1-M, MIROC-ESM-CHEM, CCSM4, CSIRO-Mk3-6-0, HadGEM2-ES, BCC-CSM1-1	1	0.07

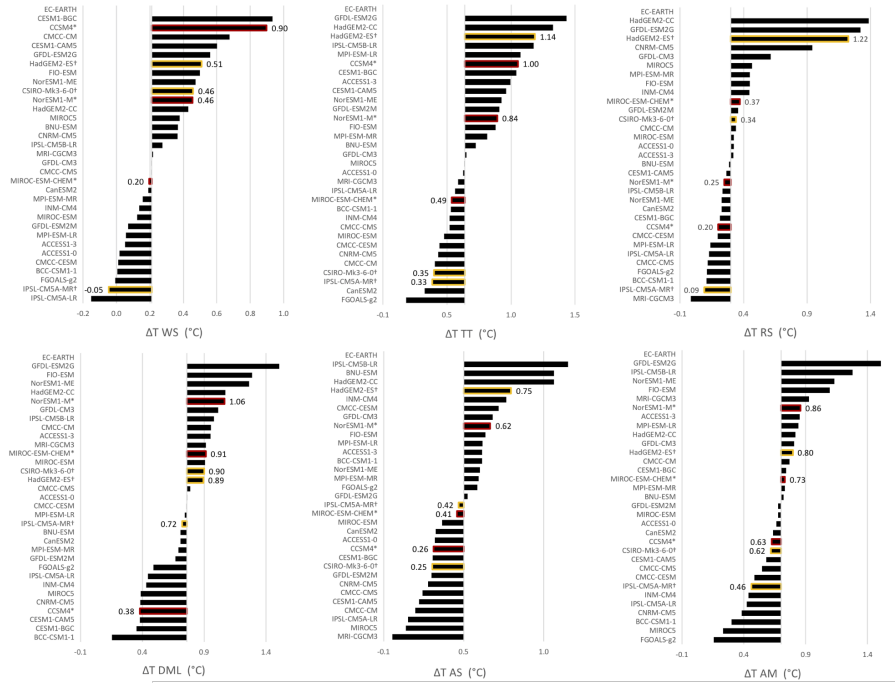


Figure B.1. Projected RCP8.5 warming for each CMIP5 model between 1980-2000 and 2080-2100 in the 6 Antarctic shelf regions (WS = Weddell Sea; TT = Totten; RS = Ross; DML = Dronning Maud Land; AS = Amundsen; AM = Amery). Labels and markers (*, †) identify models selected in the top 3 and top 6 ensembles respectively.

Table C.4. ~~Top 6 selected models for Antarctica with absolute~~ Absolute and relative occurrence of each individual model included in the Antarctica top 3 combinations (~~Tab-presented in Table C.3~~).

Models	Count	Occurrence
NorESM1-M	14	1.00
MIROC-ESM-CHEM	14	1.00
CCSM4	14	1.00
CSIRO-Mk3-6-0	14	1.00
IPSL-CM5A-MR	13	0.93
HadGEM2-ES	13	0.93
BCC-CSM1-1	2	0.14

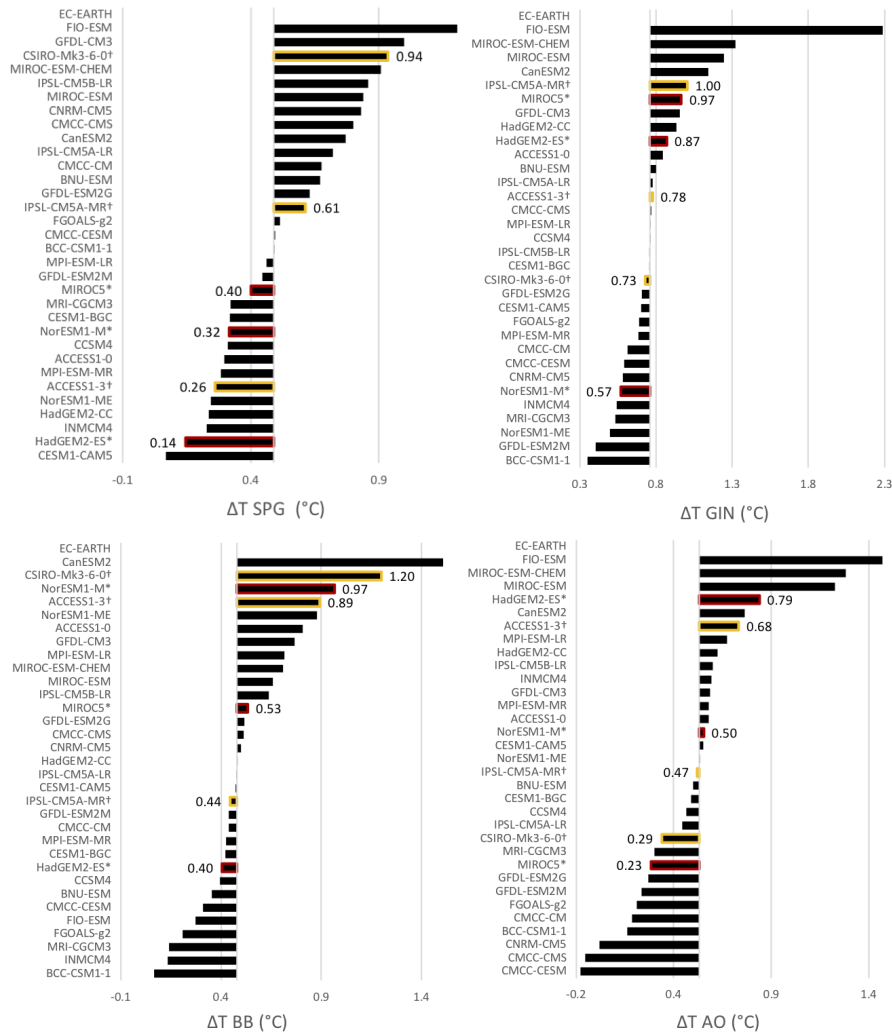


Figure B.2. Projected RCP8.5 warming for each CMIP5 model between 1980-2000 and 2080-2100 in the 4 Greenland shelf regions (SPG = Subpolar Gyre; GIN = Greenland-Iceland-Norwegian Seas; BB = Baffin Bay; AO = Arctic Ocean). Labels and markers (*, †) identify models selected in the top 3 and top 6 ensembles respectively.

Table C.5. Top 3 selected Possible model combinations for the Greenland top 3 selection, with absolute and relative frequency of occurrence in when applying the robustness test.

Model combination	Count	Occurrence
MIROC5, NorESM1-M, HadGEM2-ES	9	1.00

Table C.6. ~~Top 3 selected models for Greenland with absolute~~ Absolute and relative occurrence of each individual model included in the Greenland top 3 combinations (~~Tab. presented in Table C.5~~).

Models	Count	Occurrence
MIROC5	9	1.00
NorESM1-M	9	1.00
HadGEM2-ES	9	1.00

Table C.7. ~~Top 6 selected~~ Possible model combinations for the Greenland top 6 selection, with absolute and relative frequency of occurrence in when applying the robustness test.

Model combination	Count	Occurrence
MIROC5, IPSL-CM5A-MR, CSIRO-Mk3-6-0, NorESM1-M, HadGEM2-ES, ACCESS1-3	7	0.78
MIROC5, IPSL-CM5A-MR, CSIRO-Mk3-6-0, NorESM1-M, CCSM4, HadGEM2-ES	2	0.22

Table C.8. ~~Top 6 selected models for Greenland with absolute~~ Absolute and relative occurrence of each individual model included in the Antarctica top 3 combinations (~~Tab. presented in Table C.7~~).

Models	Count	Occurrence
MIROC5	9	1.00
HadGEM2-ES	9	1.00
NorESM1-M	9	1.00
IPSL-CM5A-MR	9	1.00
CSIRO-Mk3-6-0	9	1.00
ACCESS1-3	7	0.78
CCSM4	2	0.22