

We thank the reviewers for their careful reading and comments which helpfully contribute to the clarification and overall improvement of the manuscript.

We are addressing the comments below (in blue), as well as keeping track of the corrections to the text (in green). It should be noted that all mentions of line and figure numbers are made according to the numbering of the original manuscript.

Reviewer #1:

This paper considers a series of simulations using the iLOVECLIM model to analyze the effect of different choices of LGM boundary conditions, parameterizations, and ad-hoc modifications (to wind stress and freshwater forcing) on the LGM solution, focussing on the surface climate as well as the deep ocean overturning. Although the results are generally interesting and worthy of discussion, I think the comparison between these experiments and conclusions about which "choices" are better or more important than others need to be drawn a bit more carefully, as these various modifications aren't really comparable and some are certainly more physical than others. The paper specifically highlights the use of a parameterization to represent "sinking brines" as key to obtaining a realistic simulation of the LGM overturning circulation. However, I believe that the discussion of these results needs improvement, as elaborated below. Most importantly, I think that a PI simulations with this parameterization needs to be presented for comparison. Since the physics of the ocean have not changed between the present and LGM, the model needs to be able to reproduce the PI ocean and LGM ocean circulation with the same parameterization.

We propose here corrections on the conclusions and elements of discussion according to the reviewer's suggestions. In particular, we show the results obtained for a PI simulation with the parameterization of the sinking brines.

While we understand the reviewer's legitimate concern about the fact that our simulations are relying on modifications which are more or less physical, we would like to stress that these simulations are only sensitivity tests and do not accurately reproduce the physics of the ocean. Still, they give us the opportunity to draw conclusions about key processes, (here quoted from reviewer #2) in particular "that the sensitivity tests with winds, brines, and freshwater have more potential to influence the simulated surface properties, and particularly the ocean overturning circulation and distribution of water masses, compared to the choice of boundary conditions. This highlights the importance of informed choices in model parameterizations of processes. In particular, it further clarifies the effect of deep-water formation and convection processes for achieving a realistic representation of the glacial deep ocean and its water masses."

Our intention is certainly not to prescribe any modelling choice as key. When observing that the simulation 'P4-I brines' is the "best" one (with quotation marks, L330), we meant that it is the one that is in best agreement with geological data, given the variables analyzed in this study. We additionally qualified this simulation to be "quite crude" (L340) and we underlined it only as "one way of tackling this issue [of a simulated deep ocean circulation in disagreement with paleotracer data]" (L354). We hope that the proposed corrections will clarify this intention to the reader.

Specific comments:

-R1.1 : l. 15/16 in abstract. It would be good to clarify that you are referring to "different choices for LGM boundary conditions", or better yet "different choices for the LGM ice sheet topography" (see also comment below). After all, differences in boundary conditions between the PI and LGM are ultimately what has to explain the different circulation in the two climates.

Indeed. We have clarified this sentence following this comment and also the third comment of reviewer #2. This sentence now reads:

We investigate here the impact of a range of surface conditions in the Southern Ocean in the iLOVECLIM model, using nine simulations obtained with different LGM boundary conditions associated with the ice sheet reconstruction (e.g. changes of elevation, bathymetry, and land-sea mask), and/or modelling choices related to sea-ice export, formation of salty brines, and freshwater input.

Although the term “boundary conditions” might not be clear to a non-modeller reader, the change of boundary conditions due to the choice of different ice sheet reconstructions encapsulates changes of elevation, albedo, bathymetry, and land-sea mask. We prefer to use the term “ice sheet reconstruction” rather than “topography”, because we are concerned that the latter might be understood essentially as a change of elevation.

Note that we have noticed a few typos along the manuscript concerning the use of a hyphen in the expression “ice sheet reconstruction”, so we corrected them as well.

-R1.2 : In various places (e.g. l. 36-38) the issue of open ocean convection versus sinking along the AA slope seems to be used almost synonymously to the role of brine rejection in deep water formation, but these are rather different processes. Notably, the CCSM3 LGM simulation shows very salty AABW, clearly as a result of strong brine rejection, yet I assume AABW is still formed by open ocean convection (e.g. Shin et al. 2003, <https://doi.org/10.1029/2002GL015513> - indeed this paper should be discussed).

In order to avoid confusion between the two processes, we removed the mention to brine rejection at L. 36-38.

As reviewer #2 also asked to develop this part, we included the following in the new version of the manuscript:

Moreover, Heuzé et al. (2013) showed that, even in present-day conditions, models generally simulate inaccurate bottom water temperatures, salinities and densities. Even when they do simulate relatively accurate modern bottom water properties, they still tend to form AABW via the wrong process (namely open ocean deep convection) whereas the largest proportion of AABW currently results from formation of dense shelf waters, overflowing in the deep ocean (Orsi et al., 1999; Williams et al., 2010). While some high resolution CMIP6 models now simulate dense shelf waters, Heuzé (in review, 2020) observed no obvious export of these waters, and open ocean deep convection remains a much too widespread and frequently occurring process.

We also thank the reviewer for pointing out that the Shin et al. (2003) paper is relevant for the present study. We think a mention to this reference would fit in this part of the introduction well:

And indeed, PMIP models struggle to reproduce the glacial sea-ice extent suggested by sea-ice proxy data, and especially its seasonality (Roche et al., 2012; Goosse et al., 2013; Marzocchi and Jansen, 2017). While Ferrari et al. (2014) have shown a dynamical link between the deep ocean circulation and Antarctic sea ice, Shin et al. (2003) have highlighted the major role played by Antarctic sea ice on the glacial AMOC by quantifying the haline density flux increase at the LGM in the CCSM model. Moreover, Marzocchi and Jansen (2017) have quantitatively attributed part of the observed discrepancies of the AMOC simulated by PMIP3 models to insufficient sea-ice formation and export. Therefore, targeting sea-ice biases in models may be necessary to improve the simulated water mass distribution.

-R1.3 : Relatedly, the parameterization of sinking brines needs to be described a little more— both in terms of the formulation and its physical interpretation. I understand that this method has been published previously, but since it is key to the presented conclusion I think the reader needs to be able to interpret the results from these simulations without first reading Bouttes et al. (2010). In l. 125 it is argued that "its objective is to account for the sinking of dense water along the Antarctic continental slope", and similarly in l. 301 it is referred to as "the parameterization of the sinking of dense water along the continental slope". This gives the impression that it may be a parameterization of downslopes gravity currents, which, however, seems quite misleading. Indeed,

if I understand correctly, the parameterization simply transfers salt from brine rejection directly and locally to the bottom of the ocean (without any mixing along the way), and it is not limited to the Antarctic slope. Personally, I'll have to admit that this parameterization seems rather unphysical to me (even gravity currents are associated with lots of entrainment and detrainment as they proceed down the slope, and of course they only exist on the slope). The readers can form their own opinion, but to do so, the parameterization needs to be discussed clearly.

As the expressions “along the continental slope” may indeed induce confusion with the downslope currents, we have removed them at L.125 and L.301. Also, for the reader to understand better both the formulation of this parameterization and its physical interpretation without having to refer to Bouttes et al. (2010), we expanded the following sentences as follow:

We ran ‘P4-I brines’ using the parameterization of the sinking of brines described by Bouttes et al. (2010). The objective of this parameterization is to account for the sinking of dense water rejected during sea-ice formation. Indeed, this process is often limited by the horizontal resolution of models, as the rejected salt tends to get diluted in the surface grid cells where sea ice is forming. This parameterization allows for a fraction of the salt content of the surface grid cell to be transferred to the deepest grid cell underneath the location of sea-ice formation. As a result, the salinity and density of the bottom cells increase while the salinity and density of the surface grid cells decrease, without congruent motion of water masses. The modification of the salinity depends on the rate of sea-ice formation, as well as the chosen fraction parameter. Here the fraction was chosen at 0.8 to allow for a large effect of this sensitivity test, but the gradual effect of this parameter choice on the streamfunction is shown in Fig. S5, as well as the impact of this parameterization on the PI streamfunction (and deep water mass properties, see Fig. S6). This simple parameterization is relatively different than a downsloping current one as it is not confined to the continental slope and does not create mixing along the way of the sinking brines. While “this brine mechanism is idealized, it reflects the impact of intense Antarctic sea-ice formation during the LGM” (Bouttes et al. (2010)) on the AABW density.

-R1.4 : As discussed in the general comments above, the "P4-I brines" simulation needs to be compared to a corresponding PI simulation with the same parameterization. Ultimately the changes in the ocean circulation between the PI and LGM climates have to be attributable to differences in the boundary conditions, not different physics. It needs to be verified that the model is able to reproduce a reasonable solution for both the LGM and PI ocean with the same parameterization. (One aspect that should be paid attention to here are the T and S properties of NADW and AABW. Importantly, AABW is fresher than NADW in the modern climate, which needs to be reproduced by any model that adequately represents water mass transformation processes in the SO.)

We agree with the reviewer. We have run a ‘PI brines’ simulation with the same parameter. The streamfunction obtained is presented in Fig. 1 below. We have also added it in SI for interested readers (see Fig. S5), alongside the streamfunctions of ‘PI’, ‘P4-I’ and ‘P4-I brines’ to enable an easier comparison, as well as other simulations that the reviewer #2 mentioned (a ‘New P2 brines’ and two ‘P4-I brines’ with a different parameter). We observe that despite the large parameter (0.8) used in ‘PI brines’, the differences with ‘PI’ remain relatively small. There is a change of the Southern Ocean overturning cell (which is expected given that the parameterization is transferring salt to the deep ocean without explicitly advecting water masses) and also a small strengthening of the bottom overturning cell in the Atlantic, but the cell limits remain similar. Therefore, we expect no significant change in the water mass distribution. As a result, we think that adding the results from ‘PI brines’ as a 10th simulation in all parts of the manuscript is not critical.

As for the T and S properties of deep water masses, they are now shown in Figure S6: please refer to our response to your twelfth comment for more detail.

It can also be noted that the chosen parameter for a ‘PI brines’ simulation (and more generally the brine parameterization) has been discussed in the reviews of Bouttes et al. (2010), which can be found at: <https://cp.copernicus.org/articles/6/575/2010/cp-6-575-2010-discussion.html>.

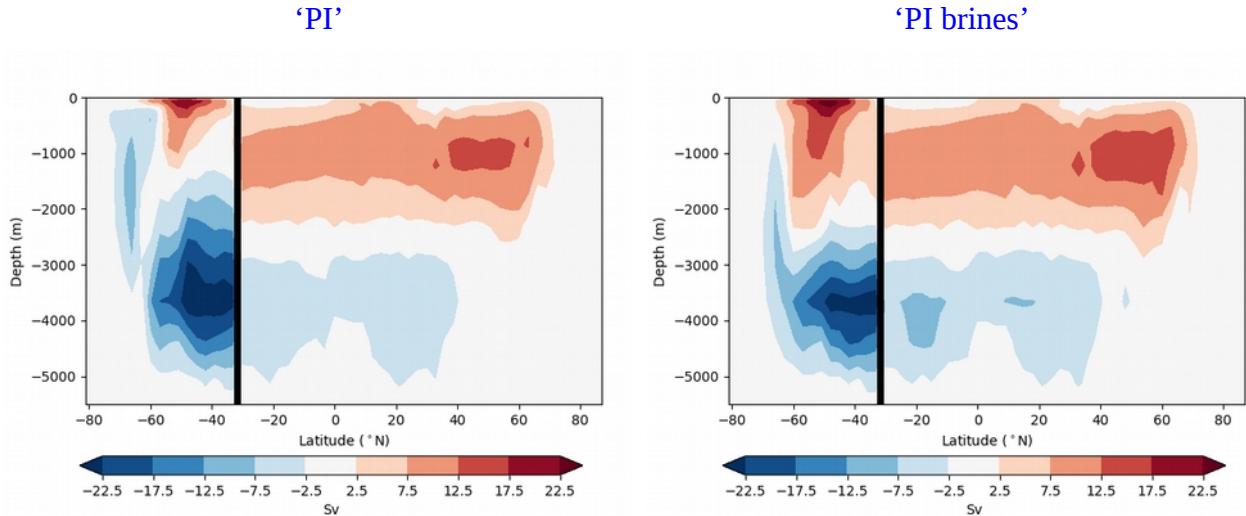


Figure 1: Streamfunctions (Sv) (as in Fig. 7) of ‘PI’ and ‘PI brines’ simulations

-R1.5 : in l. 131/132 it is argued that a quasi-equilibrium state is ensured. How is this evaluated?

We calculated the drift in time series of the global deep ocean temperature to assess when the equilibrium state is achieved – which is usually the case after 1000 years of run, though it may be a bit longer for simulations such as ‘P4-I brines’. We propose clarifying this point like this:

Each simulation has been run either 3000 or 5000 years to ensure a quasi-equilibrium state. The drift for any individual simulation is less than 2.10^{-4} °C per century for the deep ocean temperature (global mean of all oceans below 2,000 meters depth).

-R1.6 : l.161: does "surface extent" here refer to sea ice extent or still the surface of the continent? I assume the former, but please clarify.

Indeed. We have added the term “sea-ice” to avoid confusion.

-R1.7 : The error bars of 10% and 20% for winter and summer sea ice extent seem to be mostly accounting for the uncertainty in the continental margins, which is probably relatively small compared to the large uncertainty in the sea ice line. As a result, these error estimates seem very optimistic to me. Indeed this seems to be confirmed by the fact that the previous estimates of Gersonde et al. (2005) and Roche et al. (2012) fall significantly outside of this error bar. I’m not arguing that either estimate is better or worse, but simply that a larger uncertainty has to be acknowledged. I think the uncertainty range should at least encapsulate the best estimates of these major previous studies.

We here stress that both Gersonde et al. (2005) and Roche et al. (2012) used a different projection system (South Pole stereographic projection that increases distance with decreasing latitude) and subtracted the modern Antarctic ice-sheet surface to calculate the LGM winter and summer sea-ice extent instead of the LGM one. Consequently, the winter sea-ice extent of 39×10^6 km² and 43.5×10^6 km² presented in these studies was overestimated. We here used a discretization of the sea-ice line approximated on a grid with a fixed latitudinal and longitudinal spacing on a sphere and subtracted the LGM Antarctic ice sheet. We also added new control points from Benz et al. (2016)

refining the winter sea-ice extent in the Pacific sector. The recalculated area of $32.9 \times 10^6 \text{ km}^2$ is therefore much more robust. Applying a similar approach to the winter sea-ice extent in Gersonde et al. (2005) and Roche et al. (2012) would provide very similar values as the one presented here. For this reason, our error bars do not need to encapsulate previous (overestimated) sea-ice extent estimates.

We however agree that error bars of 10% and 20% seem optimistic, especially as the summer sea-ice edge is poorly constrained from the data point of view. There is a clear lack of control points and the estimation of the sea-ice line is partly done by default (we know it is South of the marine cores with no indication of sea-ice presence, but where exactly?). We could easily double these values, though 20% seems a little pessimistic for the winter sea-ice edge which is relatively well-constrained. To find a middle ground, we have chosen to change the error bar values to 15% and 30% in Fig. 5 and S3, and corrected the text accordingly.

However, it should remain clear for the reader that these values are only indicative. In fact, it is impossible to compute error bars in the usual sense, given the uncertainty of the reconstructed sea-ice edge drawn based on point data. We still wish to provide the reader with some sense of the magnitude of the error, which is why we quantified the order of magnitude of the error linked with the discretization of both this reconstructed sea-ice contour and the Antarctic continent contour at the LGM (L161-173), which adds a further element of uncertainty to the model-data comparison of the sea-ice extent. We assumed that the error associated with the sea-ice line is of the same order of magnitude as the error linked with the discretization, given the relatively coarse resolution of the CLIO model. Still, there is no statistical test which could help use quantify the error bars accurately, so the chosen values will remain questionable in any case. This is what we meant by the use of the adjective "indicative" (L161). Therefore, it would not be right to interpret how well the simulations are doing in terms of sea-ice extent based on whether they "fall inside or outside these error bars" (which is also an argument why these errors bars should not necessarily encapsulate estimates from previous studies). Note that this is an expression that we haven't used in Section 3.2, we are simply saying which simulation "falls close" to the reconstructed sea-ice extent or tend to "underestimate"/"overestimate" the sea-ice extent compared to the reconstructed one.

To stress this element further and warn the reader that these estimated error bars should not be taken at face value, we have added the adjective "indicative" in the legend of Fig. 5 as well, and modified the following sentence (L174):

Considering the order of magnitude of these alternative estimates, error bars of 15% and 30% seem reasonable. Still, these estimates are only indicative of the order of magnitude of the error.

An alternative option is, obviously, to remove all error bars. This would, however, give a false impression of certainty to the reconstructions provided.

-R1.8 : l. 220: It is argued that "the transfer of brines leads to a cooling of the Southern Ocean". Notably, however, the cooling does not occur in the regions of AABW formation but further north. There also is pronounced warming (relative to P4-I) in the North Atlantic. Do you have an explanation for these results? And does the warmer North Atlantic play a role in explaining the relatively weak and shallow AMOC in this simulation? The focus here seems to be almost entirely on the Southern Ocean, but what's the effect of the brine parameterization in the North Atlantic?

No cooling can be simulated in the regions of AABW formation due to the presence of sea ice, as the SST is at the freezing point value. The cooling is happening further north, notably in regions of upwelling and is probably due to the enhanced stratification (and decreased convection in the SO cell). We propose adding a sentence here:

We note that the transfer of salt to the bottom of the ocean leads to a cooling of the Southern Ocean ('P4-I brines', Fig. 2f), while the opposite occurs with the addition of a freshwater flux around Antarctica ('P4-I hosing', Fig. 2h). Observed in ice-free regions (i.e. where the SSTs are not necessarily at the freezing point value), this cooling is probably a consequence of the enhanced

stratification, since a well-mixed water column in upwelling regions would tend to dampen the effect of low winter surface temperatures on the SSTs.

As for the warming of the North Atlantic, it is difficult to pinpoint the exact cause of it. We only used the brine parameterization in the Southern Ocean, so it should be a consequence of the effect of this localized parameterization on the global stratification. We observe a shallower but not weaker NADW cell in ‘P4-I brines’, a denser AABW and fresher surface salinities. These effects should have consequences on ocean heat transport: we observe for example a slightly stronger North Atlantic drift. However, we have not looked deeper into it as this is not the bulk of our study. Indeed, we focused here essentially on the Southern Ocean. We have only used the brine parameterization in the Southern Hemisphere in ‘P4-I brines’, arguably because the different bathymetry affects the convection processes, but also simply to isolate the effect of a denser AABW. Though not shown in this study, we did run a simulation with the brine parameterization implemented everywhere (with the same $\text{frac}=0.8$) and have observed a similar structure of the AMOC, but with a weaker NADW cell and a slightly weaker bottom cell as well (compared to ‘P4-I brines’), which slightly improves the model-data agreement with $\delta^{13}\text{C}$ proxy data and the atmospheric CO_2 concentration (also not shown here). We think including the results from this simulation is not really telling in the context of our study.

-R1.9 : l. 214: what is the statement that “Cold P2” is not the simulation with the best overall agreement” based on? From what is shown in the paper, it seems to at least show among the best agreement in terms of SSTs. (And believing the Tierney et al. (2020) estimate it would also be the best in terms of global mean temperature.)

Believing the estimate from Tierney et al. (2020), ‘Cold P2’ is indeed best in terms of global mean surface temperature. However, considering the regional patterns of SSTs, we learn a few things. We observe in Fig. 3 (see triangles) that the simulation ‘Cold P2’ achieves a smaller RMSE for the Southern Ocean (and especially in the Atlantic and Indian sectors of the Southern Ocean, the region with the poorest model-data agreement) due to its lower SSTs. This is not surprising given the systematic warm bias we observe in summer around $40\text{--}50^\circ\text{S}$ (Fig. 4). However, looking at the mean RMSE in winter (see crosses in Fig. 3b), we see that the RMSE for ‘Cold P2’ is higher than ‘P4-I brines’ for example: despite the lower RMSE for the Southern Ocean (Atlantic and Indian sectors, see triangles), this simulation also presents a deteriorated agreement for the Atlantic Ocean (see diamonds) in winter. This suggests that a cooling of the Southern Ocean (especially the latitudes $40\text{--}50^\circ\text{S}$ of the Atlantic and Indian sectors, not so much the latitudes $\sim 60^\circ\text{S}$ of the Pacific sector) improves the model-data agreement but it is not necessarily true if this is associated with too strong a winter cooling in the North Atlantic and Arctic Ocean (see ‘Cold P2’ compared to ‘P4-I brines’ in Fig. 2). In these regions, the difference between the simulated SSTs and the SST data from MARGO Project Members (2009) can be examined in plots similar to Fig. 4 but for the Northern Hemisphere latitudes (not shown here as the focus is on the Southern Ocean).

Basically, we observe from this model-data comparison that a simulation in relatively good agreement with regional data is not necessarily the “best” simulations with respect to global estimates – and the reverse is also true. This fact is actually a limitation of the Tierney et al. (2020) data assimilation method, as their global estimate is based on the simulation with the best multi-regional agreement to data, and is not freed of potential model biases in less well-constrained regions.

Without going into too much detail, we have clarified the sentence L.214 like this:

The simulations with a colder Southern Ocean (‘Cold P2’, ‘P4-I brines’) show a better agreement with the SST data, as indicated by a smaller RMSEs computed for the Southern Ocean (see triangles in Fig. 3). However, ‘Cold P2’ is not the simulation with the lowest mean RMSE (see crosses in Fig. 3b), as it notably shows a higher RMSE in the Atlantic basin in winter (see diamonds).

-R1.10 : Fig. 7: What exactly is plotted here? Is it only the resolved Eulerian mean overturning or does this include the parameterized eddy transport associated with the GM parameterization? What matters for the transport of physical and geochemical tracers is really the isopycnal overturning (which probably does not have two counter-clockwise cells in SO). Computing the latter is admittedly more challenging and not commonly done in studies like this, but at least the GM contribution should be included.

We thank the reviewer for making us aware of this relevant issue. We checked how the streamfunction output of the model is computed: the contribution from the GM parameterization is indeed added to the Eulerian velocities. As for the isopycnal overturning, it has never been computed in the iLOVECLIM model and would be relatively difficult to do in a reasonable time considering our current workforce.

-R1.11 : From Fig. 7, it also seems that the AMOC in the PI simulation is too weak and too shallow, which should be discussed.

The AMOC in the PI simulation is indeed relatively weak and shallow, though it is not an outlier in the PMIP3/PMIP4 ensemble (see Fig. S1 and S2 of Kageyama et al. (in review, 2020)): the pre-industrial AMOC simulated by iLOVECLIM is fairly comparable to the pre-industrial AMOC of HadCM3, AWIESM2, MIROC-ESM and CNRM-CM5, and actually stronger and deeper than that of IPSL-CM5A2 (and IPSL-CM5A-LR).

This text has been included in the new version of the manuscript as follow:

The AMOC in our PI simulation is within the PMIP3/PMIP4 ensemble (see Fig. S1 and S2 of Kageyama et al. (in review, 2020)). In more details, it is fairly comparable to the pre-industrial AMOC of HadCM3, AWIESM2, MIROC-ESM and CNRM-CM5, and actually stronger and deeper than that of IPSL-CM5A2 (and IPSL-CM5A-LR).

-R1.12 : For the evaluation of deep ocean water mass properties it would be very useful to show T and S (as a function of latitude and depth).

Following this suggestion from the reviewer, we added plots of T and S distribution in the Atlantic ocean in Figure S6. These deep water mass properties were previously evaluated in the iLOVECLIM model under PI conditions in Bouttes et al. (2015) (Figures 6 and 7), using data from the World Ocean Atlas 2009. We reproduced the same type of plots for our 'PI' and 'PI brines' simulations, using the same data for comparison. We observe that using the brines parameterization (with the fairly high parameter choice of 0.8) warms the deep ocean interior by a few degrees, which deteriorates the agreement with WOA09 data. Moreover, while this parameterization has no significant effect on the temperature patterns at the subsurface, it causes a decrease of the surface salinity. At depth, although it lessens the salinity difference between the NADW and the fresher AABW, it also improves the salinity patterns in the North Atlantic mid-depths with respect to WOA09 data.

-R1.13 : l. 303-305: I don't follow the argument here about why 'P4-I wind' has a stronger Southern Ocean cell. My guess would be that the stronger wind stress over ice leads to enhanced ice export, which in turn leads to more new sea ice formation and thus brine rejection (c.f. Shin et al 2003).

The multiplication of the wind tension over ice indeed leads to enhanced sea-ice export and as a result increases the sea-ice area (Fig. 5), sea-ice formation and brine rejection at high latitudes. It also leads to sea ice piling up (increased sea-ice thickness). This was not clearly stated in the manuscript, which is why we propose the following modification:

On the other hand, the Southern Ocean cell is enhanced for 'P4-I wind', but moderately ('P4-I hosing') or strongly ('P4-I brines') suppressed for the other sensitivity tests. These results could be

due to the fact that the experimental setting of 'P4-I wind' – with the multiplication of the meridional wind stress on ice – enhances sea-ice export, which leads to an increased sea-ice formation and its consequent brine rejection (Shin et al., 2003).

-R1.14 : In Fig. 8 and throughout much of the manuscript "convection" seems to be used synonymously to the large-scale overturning circulation. However, convection can occur without a large-scale overturning and vice versa. I suggest to replace all references to convection cells with overturning cells.

We agree that such a terminology is more appropriate. We have changed it accordingly in Section 3.4 and in the captions of Fig. 8 and S5.

-R1.15 : In section 4.1 the various simulations are separated into those that amount to different choices for "boundary conditions" and "experimental setting", a separation that makes its way also into the abstract and conclusions. This separation, and the term "experimental setting" seems very vague. E.g. the assumed glacial temperature profile affects the simulation results via heat flux in or out of the glacier surface, and thus effectively also amounts to a difference in boundary conditions. In general it seems that "boundary conditions" is only used for cases with different choices for ice sheet topography, so I suggest to simply be explicit about that. As for the various other experiments, I don't see how they can be lumped into one category.

We also thought that the term "experimental setting" may be too vague, so thank you for the confirmation. After discussing the terminology, we find that this term is too general as it encapsulates both changes of boundary conditions (which are the changes of elevation, albedo, bathymetry and land-sea mask associated with the choice of the ice-sheet topography, such as GLAC-1D, ICE-6G-C or ICE-5G), and the modelling choices made in sensitivity tests, which concern either forcings (such as in 'Cold P2', 'P4-I hosing', 'P4-I wind'), or model parameter choices ('P4-I brines', as 'P4-I' would be with a chosen fraction of 0). This inadequate use of terminology is probably why the distinction of these different types of modelling choices seems vague. It should be clearer with the following specifications and corrections or the term "experimental setting":

- L. 72: Since the ice sheet reconstructions are still associated with large uncertainties, Kageyama et al. (2017) describe the common experimental design for LGM experiments in the current phase 4 of the project but let modelling groups choose from three different ice sheet reconstructions: GLAC-1D (Tarasov et al., 2012), ICE-6G-C (Peltier et al., 2015; Argus et al., 2014), or PMIP3 (Abe-Ouchi et al., 2015). To see the impact of such a choice, we have implemented in this study the boundary conditions (e.g. elevation, bathymetry, land-sea mask) associated with the first two options since these reconstructions are the most recent.

- L.54: "boundary conditions and other experimental settings"
- L.204: "the choice of boundary conditions and of the sensitivity tests"
- L. 295: "using different boundary conditions and/or forcing or model parameter choices (in the sensitivity tests)"
- L.286: "modelling choices"
- L. 299: "choice of forcings and model parameters"
- L.317: "relative impact of boundary conditions and other modelling choices (related to forcings or model parameter choices)"
- L.323: "the modelling choices made in sensitivity tests"
- L. 394: "sensitivity tests"

-R1.16 : Section 4.3.: Given the high uncertainty, particularly in the reconstructions of summer sea ice cover, I think it would be useful to provide some estimate of uncertainty for the sea ice seasonality from proxy data.

If we consider (as we first did) indicative error bars of 10% and 20% for the winter and summer sea-ice extent respectively, we get a seasonality estimate of $22.7 \times 10^6 \text{ km}^2 \pm 5.3 \times 10^6 \text{ km}^2$. Considering higher indicative error bars of 15% and 30%, we get an estimate of $22.7 \times 10^6 \text{ km}^2 \pm 8.0 \times 10^6 \text{ km}^2$, though this high uncertainty is mostly induced by the chosen error bar of 15% since the winter sea-ice extent is much larger (yet more well-constrained) than the summer one. If it was a strict estimate of the uncertainty, we would need to be more cautious with statements such as L.256-257 (“This suggests that the enhanced seasonality of the LGM Southern Ocean sea ice ($22.7 \times 10^6 \text{ km}^2$ according to our proxy reconstructions, compared to the modern seasonal range of $15.4 \times 10^6 \text{ km}^2$) is not entirely simulated by the model [...]”). However, as mentioned above, these estimates are only meant to give the reader a sense of the magnitude of the error, which is why we propose the following cautious addition:

First of all, the simulated seasonal amplitude of sea ice is too small with respect to the proxy data estimates, which suggest a sea-ice seasonality of $22.7 \times 10^6 \text{ km}^2 \pm 8.0 \times 10^6 \text{ km}^2$ based on 15% and 30% error bars on winter and summer sea-ice extent, respectively).

-R1.17 : I find the last paragraph of section 4.3 and specifically the attempt to reconcile the conflicting results between this study and Heuze et al. (2013) hard to follow and it seems very speculative. I don't think this discussion is necessary either, so I suggest removing this paragraph.

Having reflected on the previous text of the paragraph, we indeed found it difficult to follow in relation with the previous paragraphs. We therefore propose to modify it as follow, in the hope that it will be much clearer:

Identifying the origin of a bias is always a challenge. It might be an especially hard task to identify the origin of biases in the simulated sea-ice cover, considering the sheer number of feedbacks involved (Goosse et al., 2018). What can be noticed is that the simulated sea-ice seasonal cycle is affected by some of our modelling choices (increased in ‘P4-I brines’, reduced in ‘P4-I wind’). Alongside, the Southern Ocean convection is suppressed in the first sensitivity test, and enhanced in the second. In a climatological mean in our model there seems to be a link between reduced Southern Ocean convection and increased sea-ice seasonal cycle. In opposition to this observation, Heuzé et al. (2013) have underlined the fact that CMIP5 models with a large sea-ice seasonality are also the ones simulating open ocean convection over extensive areas at modern times, arguing that strong sea-ice formation could precondition the ocean for open ocean deep convection. This questions the relative importance of the different simulated mechanisms at play linking the ocean convection and the sea-ice seasonal cycle, an aspect that is present in several studies (e.g. Marshall and Speer, 2012, Behrens et al., 2016; Ma et al., 2020).

-R1.18 : Based on the issues pointed out above, I'm not sure the last sentence of the conclusion can be justified. At the least, "boundary conditions, such as the ice sheet reconstruction" should be reduced to just "the ice sheet reconstruction".

We removed the term “boundary conditions” and added “for the variables analyzed in this study” to be more accurate. Still, we would like to point out that thanks to our set of simulations, we tested the impact of different types of LGM boundary conditions: ICE-5G, ICE-6G-C and GLAC-1D differ in terms of elevation (mostly), but also albedo, land-sea mask and bathymetry, whereas it is only the ocean boundary conditions which are different in the simulations ‘Warm P2’ and ‘New P2’. As mentioned before, we are concerned that the term “ice sheet reconstruction” might be automatically associated with a change of atmosphere boundary conditions in the reader’s mind, which is why we propose the following clarification:

For the variables analyzed in this study, it would therefore seem that the correct simulation of convection processes is paramount, and far more important than the choice of ice sheet reconstruction used to implement the orography and bathymetry.

We are not implying here that the convection processes simulated (or rather, parametrized) in ‘P4-I brines’ are “correct”. This simulation is only useful to show that the intensity of the open ocean convection which is simulated in the Southern Ocean by iLOVECLIM and the large majority of models (also see Section 4.2) might be detrimental to a realistic representation of both the water mass distribution and the sea-ice seasonality. This could be of importance for other modelling groups (especially those also working on EMICs) since targeting the representation of convection processes in models may be more critical to improve the simulated glacial deep ocean circulation than spending time to implement the boundary conditions associated with different ice-sheet reconstructions in order to account for the uncertainties in the reconstructions (as suggested by Kageyama et al., 2017).

We also note that the simulation ‘New P2 brines’ (added on the second reviewer’s suggestion in Fig. S5) yields a result similar to ‘P4-I brines’ (which is also the case of a simulation ‘P4-G brines’ not shown here), which strengthens this conclusion on the secondary impact of the choice of boundary conditions at the LGM (at least for the model resolution and variables examined in this study).

Reviewer #2:

In this paper, Lhardy et al. use an intermediate-complexity model to evaluate the response of the simulated glacial ocean circulation to Southern Ocean surface ocean temperature and sea-ice conditions. They achieve differences in these surface conditions by running simulations with various options for glacial climate boundary conditions from the Paleo Model Intercomparison Project (PMIP), and by simulations where they change wind (i.e. sea-ice export) conditions, formation of salty brines, and freshwater input in the Southern Ocean. They find that the sensitivity tests with winds, brines, and freshwater have more potential to influence the simulated surface properties, and particularly the ocean overturning circulation and distribution of water masses, compared to the choice of boundary conditions. This highlights the importance of informed choices in model parameterizations of processes. In particular, it further clarifies the effect of deep-water formation and convection processes for achieving a realistic representation of the glacial deep ocean and its water masses. This is of importance for our ability to understand, and simulate, glacial ocean storage of CO₂, and to improve our models. As this category of models is commonly used for paleo simulations, this study is particularly educational for modelling groups using models of similar resolution and complexity, especially those participating in PMIP.

The experimental design is overall sound, though adding a PI state that uses the brines parameterization would be beneficial, to test the effect of ‘a better representation of deep-water formation’ in the modern ocean. As the parameterization does not change the amount of sea ice, only its effect on water mass properties and circulation, the choice of 0.8 as the scaling should not have to change between climate states. The amount of brines, and thus their influence on the ocean properties, should decrease in a warmer climate due to the reduction in sea-ice. It would also be useful to test the brines parameterization together with the PMIP2 boundary conditions, to strengthen some of the conclusions about the role of boundary conditions, but I leave it up to the authors to decide if this is feasible.

Overall, the paper is well written, with a generally clear structure and informative figures. However, some clarifications, motivations of choices, and rephrasings are advisable prior to publication. I therefore suggest the following revisions.

We thank the reviewer for these overall positive comments and appreciate the proofreading and meticulous suggestions provided which made our revision work easier. Following the two reviewers’ comments and thanks to the relatively low computation time of iLOVECLIM, we were able to performed additional simulations with PI and PMIP2 boundary conditions with the brine parameterization (under the same model version) to answer these specific questions. We also propose corrections and elements of clarification to the points raised here.

Please note that we disagree with part of the following statement: “As the parameterization does not change the amount of sea ice, only its effect on water mass properties and circulation, the choice of 0.8 as the scaling should not have to change between climate states. The amount of brines, and thus their influence on the ocean properties, should decrease in a warmer climate due to the reduction in sea-ice”. We observe that the use of the brine parameterization has an effect on sea ice, increasing both the sea-ice extent and its seasonal amplitude, and therefore further stimulating the transfer of salt to the bottom ocean. In addition, while the reduction of sea ice in a warmer climate indeed decreases the amount of brines transferred to the bottom of the ocean, a change of scaling parameter could be argued due to the impact of the higher sea level on the formation of dense water overflows on the continental shelves, as the sheer surface of these shelves is not the same for the PI. While these reasons could potentially justify the use of a different scaling parameter (or the use of a more complex parameterization), we did keep a parameter of 0.8 in order to make the PI simulation comparable.

Finally, although we mentioned that “The parameterization of the sinking of brines is the only experimental setting we used which accommodates a better representation of both the surface conditions and the deep ocean circulation (L408-409)” (we corrected a typo here, we meant “circulation” and not “distribution”), we are not sure of where we stated that this parameterization allows for “a better representation of deep-water formation”. Technically this is not true. As reviewer #1 pointed out, the use of such a parameterization does not improve the model representation of the physics of the ocean. Our objective remains simply to see the impacts of a modelling choice which accounts for a physical process which is not well represented in relatively coarse resolution models due to its subgrid scale.

Abstract

R2.A1 : P. 1, L. 5 and L. 12: ‘proxy data’: please specify ‘proxy records of . . .’

R2.A2 : P. 1, L. 7: ‘with respect to data’ – data in this context is very unspecific. I would suggest ‘paleoproxy data’

We have adopted the suggested modifications.

R2.A3 : P. 1, L. 9: I suggest replacing the vague descriptions ‘different modelling choices and/or boundary conditions’ for something more specific, e.g. by rephrasing to ‘[. . .] different boundary conditions for climate and ice sheets, and choices for sea-ice export, formation of salty brines, and freshwater input [. . .]’

Following this suggestion and the first comment of reviewer #1, we have modified this sentence which now reads:

We investigate here the impact of a range of surface conditions in the Southern Ocean in the iLOVECLIM model, using nine simulations obtained with different LGM boundary conditions associated with the ice sheet reconstruction, and/or modelling choices related to sea-ice export, formation of salty brines, and freshwater input.

Introduction

R2.B1 : P. 2, L. 25: Consider adding a reference to Galbraith and de Lavergne (2019), see also my comment for the Discussion section. Galbraith, E., de Lavergne, C. (2019). Response of a comprehensive climate model to a broad range of external forcings: relevance for deep ocean ventilation and the development of late Cenozoic ice ages. Climate Dynamics, 52(1-2), 653-679.

We thank the reviewer for suggesting adding a reference to this article. We have done so in the Discussion section. However, we think a reference to this article at P.2 L. 25 (“Rearrangement of

water masses explains part of past changes in the carbon storage capacity of the oceans (Buchanan et al., 2016; Khatiwala et al., 2019; Yu et al., 2016), which stresses the importance of correctly simulating the processes affecting the deep ocean circulation.”) might not fit well with the other articles cited here. Although it is commonly accepted that the rearrangement of water masses at the LGM has had an influence on the carbon storage capacity of the ocean, the contribution of this process relative to others (changes of the biological pump efficiency, sea ice, and CO₂ solubility) is still much debated. In particular, Khatiwala et al. (2019) have underlined a dominant role of the air-sea disequilibrium (and a relative small role of the ocean circulation) on the glacial CO₂ drawdown. Hence the term “part of”. Since Galbraith and de Lavergne (2019) did not assess the contribution of these different processes (and they could not, considering that their “biogeochemical model does not include any air-sea effect of the variable pCO₂” and “simulates changes in dissolved nutrients and gases that are caused by the changes in ocean circulation only [...]”), we felt (but may be mistaken) that this reference would not match what is implied here by the term “part of”.

R2.B2 : P. 2, L. 53: ‘three sensitivity tests’ - Please mention what these are e.g. ‘three sensitivity tests of Southern Ocean conditions for sea-ice export, formation of brines, and freshwater input.’ (see also suggestion above for the Abstract)

We have followed this suggestion.

R2.B3 : P. 2, L. 36-38: Model representation of Southern Ocean deep-water formation is rather central for the conclusions of this paper. I would suggest clarifying the reasons for why deep water is formed by the wrong process in most models.

Although we are unsure of the technical reasons why models are convecting too easily, we can elaborate further on Heuzé (in review, 2020) conclusions of a “too deep, too often, over too large an area” open ocean convection. Given this suggestion and the second comment of reviewer #1, we propose the following sentence:

Moreover, Heuzé et al. (2013) showed that, even in present-day conditions, models generally simulate inaccurate bottom water temperatures, salinities and densities. Even when they do simulate relatively accurate modern bottom water properties, they still tend to form AABW via the wrong process (namely open ocean deep convection) whereas the largest proportion of AABW currently results from formation of dense shelf waters, overflowing in the deep ocean (Orsi et al., 1999; Williams et al., 2010). While some high resolution CMIP6 models now simulate dense shelf waters, Heuzé (in review, 2020) observed no obvious export of these waters, and open ocean deep convection remains a much too widespread and frequently occurring process.

Methods

R2.C1 : P. 3, L. 67: There is no mention of how this sea-ice component differs from those in other PMIP models, and how the model representation of sea ice potentially impacts the results. As a non-expert on sea-ice modules, I would have liked to see a sentence or two that discusses this.

Goosse et al. (2013) (<http://dx.doi.org/10.1016/j.quascirev.2013.03.011>) have reviewed all the sea-ice components used in PMIP models, their table 1 is in particular informative regarding this remark. The sea-ice component in iLOVECLIM simulates a visco-plastic rheology but no ice thickness distribution, and is quite classic compared to the rest of the PMIP models. Basically, this component simulates the basic processes (thermodynamics and dynamics) but in a relatively simple way, far from the complexity of the more recently developed sea-ice component (such as LIM3.6 in Rousset et al. (2015), <https://doi.org/10.5194/gmd-8-2991-2015>). We propose adding the following mention:

It also includes a thermodynamic-dynamic sea-ice component described by Fichefet and Morales Maqueda (1997). This component simulates a visco-plastic rheology but no sea-ice thickness distribution, which is relatively classic compared to other PMIP models (see Table 1 of Goosse et al., 2013) but far from the complexity of more recently developed sea-ice components (Rousset et al., 2015).

R2.C2 : P. 3. Section 2.2: I feel that it might be clearer if this section is amended to be ‘The PMIP boundary conditions and their implementation’ and thus to include descriptions of the PMIP2 boundary conditions and how they differ from PMIP4 (see specific comment for L. 73-76 below for an example)

We have done as suggested (also see our response for the next comment).

R2.C3 : P. 3, L. 73-76: It is clearly stated (much) later in the paper (P. 10, L. 318) that GLAC-1D and ICE-6G-C are the main recommendations of Kageyama et al. (2017) among multiple options. From the current phrasing here on P. 3, it is not clear why the PMIP3 option is excluded from the present study. It would be clearer if the phrasing were more similar to that on P. 10. In addition, there is no introduction to the ICE-5G option (presented on P. 4, L. 113-115), as it is part of the PMIP2 boundary conditions (see previous comment).

Indeed. We propose the following clarification, while moving up two references (Roche et al, 2007 and Peltier, 2004):

Since the ice sheet reconstructions are still associated with large uncertainties, Kageyama et al. (2017) describe the common experimental design for LGM experiments in the current phase 4 of the project but let modelling groups choose from three different ice sheet reconstructions: GLAC-1D (Tarasov et al., 2012), ICE-6G-C (Peltier et al., 2015; Argus et al., 2014), or PMIP3 (Abe-Ouchi et al., 2015). To see the impact of such a choice, we have implemented in this study the boundary conditions (e.g. elevation, bathymetry, land-sea mask) associated with the first two options since these reconstructions are the most recent. We have also considered the results obtained with the previous LGM version of the model (PMIP2) described in Roche et al. (2007), which was generated with the boundary conditions associated with ICE-5G (Peltier, 2004), a previous reconstruction with notably higher elevation of the Northern Hemisphere ice sheets.

R2.C4 : P. 4, Section 2.3: This section is very technical and not necessarily relevant to the average reader. I suggest moving it to an appendix, or include it with the rest of the description of the bathymetry generation method in the SI.

Following your advice, we have moved this section to an appendix. To keep it near the rest of the description of the bathymetry generation method, we have also moved this part to an appendix section. As a result, only figures remain in the supplementary information. We have also changed the references to these appendices and the section numbers accordingly.

R2.C5 : P. 4, Section 2.4: Please add a PI state that uses the brines parameterization, to test the effect of ‘a better representation of deep-water formation’ in the modern ocean.

Please refer to our response to the fourth comment of reviewer #1. We have added an additional figure (Fig. S5) to the SI to show to the interested readers the relatively small impact of this parameterization on the PI Atlantic streamfunction.

R2.C6 : P. 5, L. 122: Please specify why P4-I is selected as the reference LGM state over P4-G (see also comment for P. 7, L. 195)

We added a sentence here:

We added to this set three sensitivity tests. The boundary conditions associated with ICE-6C-G were arbitrarily chosen as standard in these tests, which is why the simulation ‘P4-I’ is considered as a LGM reference in the following sections. Sensitivity tests using the simulation ‘P4-G’ as reference (i.e. GLAC-1D boundary conditions) yield fairly similar results (not shown here).

R2.C7 : P. 5, L. 125: ‘a chosen fraction (here 0.8)’ – Please specify how this choice is made and how a different choice might impact the results (see also comment for Discussion P. 11, L. 338-342)

We wanted a large effect of the brine parameterization, which is why we chose a significant fraction without it being too unrealistic (e.g. 1). The fraction values which can be considered realistic have been largely discussed before in Bouttes et al. (2010) and in the associated peer review discussion. We have also explored different parameter values in sensitivity tests (fully comparable to the simulations in our paper: same restart and same model version), and the streamfunction obtained with $\text{frac} = 0.4$ and $\text{frac} = 0.6$ are now shown in Fig. S5. Most of the change occurs between $\text{frac} = 0.4$ and $\text{frac} = 0.6$, though a further weakening of the NADW cell and strengthening of the bottom cell can be observed between $\text{frac} = 0.6$ and $\text{frac} = 0.8$. Since the interested reader can now examine the impact of such a choice on the streamfunction in the SI, we suggest this small addition:

The modification of the salinity depends on the rate of sea-ice formation, as well as the chosen fraction parameter. Here the fraction was chosen at 0.8 to allow for a large effect of this sensitivity test, but the gradual effect of this parameter choice on the streamfunction is shown in Fig. S5, as well as the impact of this parameterization on the PI streamfunction (and deep water mass properties, see Fig. S6).

R2.C8 : P. 5, lines 136-139. 1) I find this paragraph to be phrased in a confusing way. I suggest separating the descriptions of LGM and PI data. 2) The MARGO Project Members reconstruct the LGM sea-surface temperatures. Please explain briefly why there is a lack of data for the Southern Ocean in austral winter. On P. 8, L. 219, you say that it is due to an extensive sea-ice cover, but coring that is done in the summer will still provide sediments from past winters, so it should be clarified why the winter sea-ice cover is a problem.

While moving the mention of the PI data used at the end of the paragraph in a separate sentence, we propose removing the confusing mention “due to an extensive sea-ice cover” and adding a few relevant details here:

The simulated surface conditions are first compared with the LGM sea-surface temperatures reconstructed by MARGO Project Members (2009). Thanks to the use of multiple proxies (diatoms, radiolaria, dinoflagellates, foraminifera, Mg/Ca, and alkenones), this dataset, combining 696 individual records, provides a synthesis of our knowledge of the LGM ocean surface temperature. However, it should be noted that most proxies are calibrated against summer SST (Esper et al., 2014; Cortese and Prebble, 2015) or annual SST (Sikes et al., 1997; Prahl et al., 2000). Only planktonic foraminifera allow for the estimation of winter SST (Howard and Prell, 1992) but their growth is hampered, and restricted to a couple of species, south of the Polar Front (Bé and Hutson, 1977). As such, there are only few winter SST estimates to compare with the simulated ones.

As for the model-data comparison of the PI SSTs, we relied on the modern WOA98 data (World Ocean Atlas, 1998) since it is the one used by MARGO Project Members (2009).

The additional references have also been listed in the Reference section of the paper.

R2.C9 : P. 5, line 137. You say here that you compare the PI simulation to World Ocean Atlas data. Please specify which version of the WOA data that is used, and if you are indeed using the WOA98, explain why you are not using the most recent version. I suspect it is because the MARGO Project Members are using WOA98, but if so, this needs to be stated clearly. If you are using a more recent

version of WOA, please cite the appropriate publications for each variable. Also, according to the figures, the PI simulation is compared to MARGO data (see comment for e.g. Fig. 4)

We are here using WOA98 for consistency with MARGO Project Members (2009) for their modern dataset. WOA98 was chosen by MARGO Project Members (2009) because the core-top data represent average conditions (decades to centuries depending on the sedimentation rates) of sub-modern conditions (generally younger than 2000 years old) that might not be reflected by the Anthropocene climate. This is now clearly stated in the manuscript that we are using WOA98 for consistency between PI and LGM data:

As for the model-data comparison of the PI SSTs, we relied on the modern WOA98 data (World Ocean Atlas, 1998) since it is the one used by MARGO Project Members (2009).

R2.C10 : P. 6, L. 160: Please mention what causes this notable difference in surface area.

We have completed this sentence as follows:

To put this value into perspective, the modern Antarctic continent has a surface area of 13.9×10^6 km² (Fretwell et al., 2013), due to a smaller areal extension of the Antarctic ice sheet and a higher sea level.

R2.C11 : P. 6, L. 161-162: ‘For the indicative error in the surface extent computed, we kept the respective values of 10

As this sentence is weirdly cut in the posted referee comment, we are not sure what the reviewer meant. But we have developed an answer to the seventh comment of reviewer #1 which concerns the chosen error bar values, maybe this provides some response to this comment as well?

Results

R2.D1 : P. 6, L. 178, Section 3: ‘Methods’- Should be ‘Results’

Indeed! This has been corrected accordingly.

R2.D2 : P. 6, L. 180: ‘Cold P2 is too cold’ – but it is well comparable to the more recent estimate by Tierney et al. (2020) mentioned later in the paragraph. I feel like this should be mentioned in the discussion of that paper and the fact that iLOVECLIM generally simulates more modest SAT anomalies (P. 7, L. 187-192), as this experiment design is an example of when the model actually achieves a more extreme anomaly.

While ‘Cold P2’ is too cold with respect to Annan and Hargreaves (2013), ‘Cold P2’ is indeed comparable to Tierney et al. (2020). However, for the reason mentioned in our response to reviewer #1 (see ninth comment), we prefer to remain cautious, especially as Tierney et al. (2020) estimate is lower than previous estimates and relies on data assimilation with the coldest model of the PMIP4 ensemble (Kageyama et al., in review, 2020). As the global mean SAT anomaly is a weighted averaged of local measurements, its reconstruction for the LGM is still debated. We have not elaborated further on the topic of Tierney et al. (2020) estimate in the discussion section than what is already mentioned L. 190-192, since the global mean SAT anomaly is not central in our study. However, a mention to this ‘Cold P2’ experimental design in the discussion actually strengthens our conclusion about the relative impact of boundary conditions and other modelling choices, so we thank the reviewer for pointing this out and we propose adding a sentence L.324:

In contrast, the choice of experimental setting can cause much larger differences (e.g. between ‘Cold P2’ and ‘Warm P2’, or ‘P4-I’ and ‘P4-I brines’, or ‘P4-I’ and ‘P4-I hosing’). In particular, the differences between ‘Cold P2’ and ‘Warm P2’ suggest that, while iLOVECLIM generally simulates

a more modest global SAT anomaly than other PMIP4 models (Kageyama et al., in review, 2020), modelling choices related to the glacial temperature profiles used in the radiative code can induce a very significant change.

R2.D3 : P. 6-9, Section 3.2-3.3: The model-data analysis in these two sections could gain from a comparison of model skill (M) as described by Watterson (1996). This allows an evaluation of overall model-data agreement (patterns and point-to-point agreement), globally as well as on a basin level, and easier comparison between ensemble members and time periods. It would quantify statements such as those made on P. 11, L. 331-333. Watterson, I. G. (1996). Non-dimensional measures of climate model performance. International Journal of Climatology: A Journal of the Royal Meteorological Society, 16(4), 379-391.

We are thankful for this suggestion of a metric that we did not know about. We have looked into it, in particular in comparison to other methodologies (as in Jackson et al., 2019, <https://doi.org/10.1016/j.envsoft.2019.05.001>). From our reading and our limited understanding of the topic, it does not seem obvious why this metric should be better than another one (cf. Jackson et al., 2019, their Table 3). Given the large amount of work required to implement this methodology and the relatively small impact it could have on our results, we feel that it may not be necessary at this point. We have however contacted colleagues more directly involved in CMIP model data-model comparison to check what the accepted methods are in that community. From an initial feedback received, the M skill score does not seem to be widely used. We will continue to gather more information on this topic to better understand the potential of such a methodology (or alternative ones) for future work.

R2.D4 : P. 7, L. 195: ‘the reference LGM simulation P4-I’ – it is never mentioned in the Methods why this simulation is chosen as the reference over P4-G. This should be clarified. Is this choice likely to influence the results of the sensitivity tests, and if so, how?

We have chosen to clarify this point in Methods rather than in L.195, so please refer to the answer to your comment on P. 5, L. 122 for detail.

R2.D5 : P. 7, L. 197-201: In Fig. 2, Southern Ocean SST anomalies in P4-G show a similar pattern as P4-I brines (with the exception of the mid-to-eastern Indian Ocean sector of the Southern Ocean). If you have an idea for why this is, it could be interesting to mention.

It does not seem to us that SST anomalies of ‘P4-G’ and ‘P4-I brines’ have a similar pattern. For example, the warm equatorial Atlantic and cold eastern Indian in ‘P4-I brines’ are not evidenced in ‘P4-G’. The absolute values of the anomalies are also quite different between the two simulations, who have a very different experimental design. However, we note the positive anomalies south of Africa, which can be related to the southward shift of the ACC we observed in the two simulations with respect to ‘P4-I’. Although this is not the core of the paper, we added a sentence to mention this (without including figures displaying the surface currents):

The differences between ‘P4-G’ and ‘P4-I’ are small (Fig. 2e), with the exception of the eastern Atlantic and western Indian sectors of the Southern Ocean, south of the African continent, where ‘P4-G’ displays warmer SSTs. This positive anomaly is related to a southward shift of the Antarctic Circumpolar Current.

R2.D6 : P. 8, L. 237-239: Please be clear also on how sea ice area is defined.

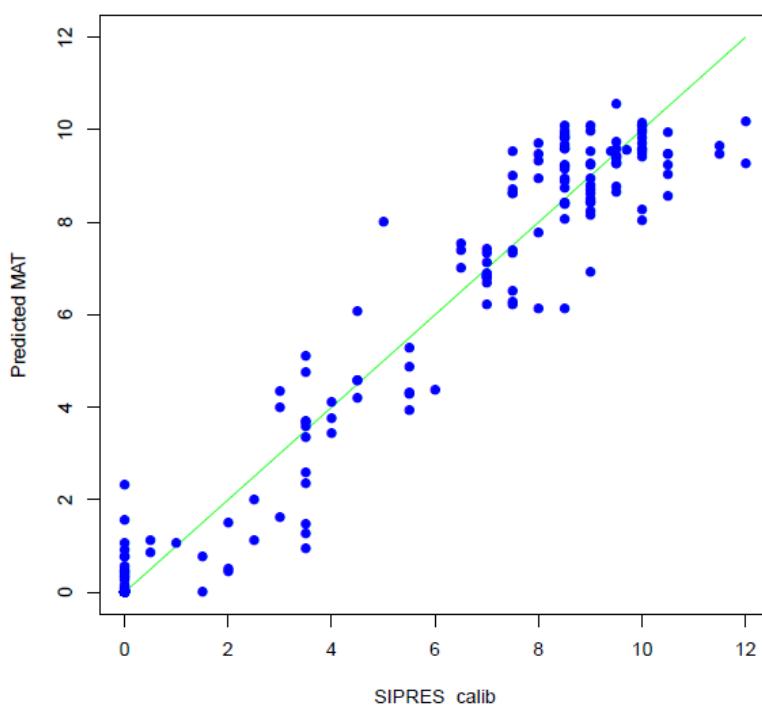
We have added the following definition:

Only the sea-ice extent – defined as the surface with a sea-ice concentration over 15% (by convention, see the US National Snow and Ice Data Center website) – is strictly comparable to our

data estimates, though we choose to present in Fig. 5 both the sea-ice extent (defined as the total surface between the northernmost 15% concentration limit and the Antarctic continent) and area (defined as the sea-ice concentration times the area of the grid cell for all ocean cells south of the equator).

R2.D7 : P. 8, L. 241: Minimum and maximum sea-ice extent - Is data available to use this method to compute corresponding numbers for the PI/modern day, to evaluate how these numbers compare to other estimates (e.g. Parkinson and Cavalieri, 2012, that is used here)? This could clarify whether the method gives rise to any systematic bias, and how well the model agrees with the different sea-ice extent estimates for the different time periods.

Winter and summer sea-ice limits were estimated through the combination of several proxies, both qualitative (relative abundances of sea-ice linked diatoms; Gersonde and Zielinski, 2000) and quantitative (Crosta et al., 2004; Esper et al., 2014). Quantitative estimates of sea ice are provided through diatom-based transfer functions that are validated on the modern model. More precisely, the modern database (composed on one hand of diatom assemblages in modern core-tops and on the other hand of modern winter sea-ice concentration or modern sea-ice duration for each modern core-top) is run onto itself. For the modern analog technique (MAT), the transfer function selects the five most similar core-tops to each modern core-top and estimates a modern sea-ice duration for each modern core-top. It allows a direct comparison of the estimated parameter (here sea-ice duration) with the measured modern sea-ice duration (here calculated from monthly sea-ice concentration at each core-top; Crosta et al., 2004). The plot below shows that the MAT adequately reconstructs the modern distribution of sea-ice duration with no obvious regional bias as the scatter along the 1:1 line (green) is pretty constant over the whole sea-ice range (0-11 months per year). Similar results are obtained with winter sea-ice concentrations (Esper et al., 2014). Diatom-based transfer functions generally reconstruct the modern model with a root mean square error of prediction of 10% for the winter sea-ice concentration (Esper et al., 2014) and 1 month per year for sea-ice duration (Crosta et al., 2004).



Biplot of the estimated sea-ice duration (PredictedMAT) in the 257 core-tops composing the diatom modern database vs the observed sea-ice duration in the numerical atlas (SIPRES calib) at the location of the 257 core-tops composing the modern database.

R2.D8 : P.8, L. 251: The background for the previous comment is the statement here that ‘the sea-ice extent of most simulations falls close to the reconstructed winter sea-ice extent’. To me, this fact seems to suggest that your maximum reconstructed extent might be an underestimation, given that most of the simulations are on the lowest end of the reference interval for glacial cooling by Annan and Hargreaves (2013).

While it is true that we have to be cautious with such statements given the uncertainty of the reconstructed sea-ice extents, the two following observations (the relative closeness of the simulated sea-ice extents to the winter reconstruction yet the modest global SAT anomaly of most simulations) are not necessarily inconsistent since they are on a different spatial scale. Compensation of errors may occur, for example: given the circular sea-ice distribution we observe, we think that a realistic sea-ice extent can be achieved while overestimating sea ice in the Pacific sector yet underestimating the sea-ice extent in the Atlantic sector. On a different scale, the regional biases in the Southern Ocean SSTs do not necessarily have to be consistent with a warm bias on the global scale: the modest global cooling observed in most simulations (with respect to the glacial cooling estimated by Annan and Hargreaves (2013) - and even more modest with respect to the one estimated by Tierney et al. (2020)) do not only reflect biases in the Southern high latitudes. Moreover, it should be noted that SST and sea-ice biases are not fully comparable as the tuning parameters used in the sea-ice model may affect their relationship.

R2.D9 : P. 10, L. 294: ‘paleotracer data’ – It would be helpful to remind the reader of the relevant references.

We have provided a few selected references here:

As Otto-Bliesner et al. (2007) have shown, iLOVECLIM is among the models which simulate a very strong glacial NADW cell at the expense of the bottom cell (as is also the case here for almost all experimental settings, see Fig. 7b,c,d,e,f,h,i), a response which is not consistent with the shallower glacial NADW and the more voluminous AABW inferred from paleotracer data (Curry and Oppo, 2005; Howe et al., 2016; Böhm et al., 2015; Lynch-Stieglitz et al., 2007).

R2.D10 : P. 10, L. 296-298: Do you have an explanation for why the enhancement of the bottom convection cell occurs as a response to the change in ice sheet boundary conditions?

Given that the surface conditions we observed for ‘P4-I’ and ‘P4-G’ are fairly similar, we think that the relatively small difference in the streamfunction may be linked to differences in land-sea mask and/or bathymetry around the coast of the Antarctic continent, which are significant in this region of large mixed layer depth (MLD). The simulation ‘P4-G’ tends to show deeper MLD maximums over this region compared to ‘P4-I’ (not shown here). As this remains hypothetical and would require including more figures, we haven’t made any modification to the text.

R2.D11 : P. 10, L. 298: ‘the simulation associated with GLAC-1D (compared to ICE-6G-C)’ – Here, it would be helpful to specify which one of P4-G and P4-I uses which ice sheets

The sentence now reads:

We also notice differences between the ‘P4-G’ and ‘P4-I’ streamfunctions, with a slight enhancement of the bottom overturning cell in the ‘P4-G’ simulation associated with GLAC-1D (compared to the ‘P4-I’ simulation with ICE-6G-C) [...].

R2.D12 : P. 10, L. 311-312: ‘showing that simulations with a colder Southern Ocean tend to be associated with a stronger Southern Ocean cell, a weaker bottom cell and a more intense NADW cell’ - Does stronger/weaker in this sense also refer to the volume occupied by the cell (i.e. depth of the water mass boundary between the bottom cell and the NADW cell)? I get this impression when I read about the results for the ‘P4-I brines’ simulation. If so, it should be pointed out that proxy

records conflict with this result, as they show a colder Southern Ocean simultaneously with a shallower NADW cell and a more expanded bottom cell. This is very well summarized in the Conclusions section.

Here the adjectives stronger/weaker refer to the maximum intensity of the overturning cells and not to the volume occupied by the cell. We thought that computing the latter value would not be informative due to the lack of a bottom cell in the Atlantic basin for most simulations except for 'PI' and 'P4-I brines'. Though related, these two variables (the intensity of the overturning and the volume of the cell) are not the same, as we can observe on the 'PI' and 'P4-I brines' Atlantic streamfunctions (the volumes are relatively close, but the NADW and bottom cells are more intense in simulation 'P4-I brines'). However, in the context of this paragraph, 'P4-I brines' is actually not a good example, since the parameterization used "artificially" impacts the density of water masses (which is why we are using a subset of simulations for Fig. 8 instead of all of them, see Fig. S5). When we simulate different surface conditions in the Southern Ocean "naturally", thanks to the use of different boundary conditions, we observe this relationship between a colder SST, a stronger Southern Ocean cell, a weaker bottom cell and a stronger NADW cell. A cooling of the Southern Ocean seems associated with a deepening of the AABW/NADW boundary. And indeed, while this relationship remains true, we do have a conflict with proxy records as the simulations cannot accommodate both a better agreement with SST data in the Southern Ocean, and a better agreement with paleoproxy data which indicate a shallower NADW.

While we don't want to go into too much detail, we could add a more conclusive sentence:

The correlation coefficients R are very significant (with $|R| \geq 0.83$ for all plots), showing that simulations with a colder Southern Ocean tend to be associated with a stronger Southern Ocean cell, a weaker bottom cell and a more intense NADW cell. While this relationship holds, modelling choices yielding colder SST in the Southern Ocean (thus in better agreement with the data) do not lead to more realistic water mass distributions. Instead, a Southern Ocean cooling seems associated with an intensification of the open ocean convection, with a negative effect on stratification.

Discussion

R2.E1 : P. 10, Section 4.1: Important aspects of the effect of boundary conditions, modelling choices, and vertical mixing on LGM simulations are all discussed in Galbraith and de Lavergne (2019). I suggest mentioning the findings of this publication somewhere in sections 4.1-4.3.

We thank the reviewer for the advice. Since there are a lot of findings in this article, we chose one which seems (arguably) the most relevant to provide insights (and hindsight) to our own results. We propose adding this sentence at the end of section 4.1:

It is therefore particularly important to investigate and understand the origin of these biases, while different ice sheet reconstructions have a relatively smaller impact and may not all be implemented during the PMIP4 exercise. Nonetheless, it should be noted that Galbraith and de Lavergne (2019) have investigated the effects of a broader range of forcings (greenhouse gas concentrations and orbital parameters in addition to changes in ice sheet size) on the deep water masses and they notably highlighted the nonlinear responses of their volume to varying forcings (e.g. with different global temperatures). Therefore, the choice of ice sheet reconstruction could potentially yield more significant differences in deep ocean circulation under different time periods or simulated global temperature.

R2.E2 : P. 11, L. 323-324, and 326-327: Based on the remark on P. 9, L. 282, that all the simulations show similar biases in seasonal and regional patterns, could you give examples of sensitivity tests that might show somewhat different biases, or do you think this is too much of a persistent characteristic of the model (if so, why)?

Considering the relatively coarse horizontal resolution of the iLOVECLIM model ($3^\circ \times 3^\circ$ for the ocean model CLIO), we do think that biases in the SST gradients (such as the one observed at 40-50°S in the Southern Ocean) are likely to remain an issue, especially as strong fronts (polar and subantarctic) are located in this region. This being said, we can still imagine sensitivity tests with a potential impact on the observed biases in seasonal and regional patterns, such as playing with the Southern Ocean westerly winds (strength, position, or effect on Ekman transport), or with the bathymetric constraint of Drake Passage: such experiments affecting winds, currents, and convection processes may impact the observed bias in surface conditions, such as the lack of interbasin contrast. In fact, we have run sensitivity tests in which we arbitrarily modified the strength and the position of the westerly winds and found a relationship between the intensity of the Southern Ocean overturning cell and the sea-ice seasonality.

R2.E3 : P. 11, L. 338-342: It would be advisable to mention the choice of the fraction 0.8 and how it would potentially affect the results if this was chosen differently (see also comment for Methods P. 5, L. 125)

We mentioned again that a model parameter choice is involved in this test by simply adding: Though legitimate, this parametrization is quite crude: a fraction (here chosen at 0.8) of the salt content of the surface grid cells is directly transferred to the deepest grid cell beneath them, without explicitly computing the convection .

As for the effects of such a choice, as shown in Fig. S5 and mentioned in our response to your comment for P. 5, L. 125, the AMOC starts to look more realistic at around $\text{frac} = 0.6$. However, if we analyze the model-data agreement with $\delta^{13}\text{C}$ data (not shown in this study), the additional improvement of using $\text{frac} = 0.8$ is significant. Note that the reference to Figure S5 was added in the Methods rather than in the main text.

R2.E4 : P. 11, L. 343-344: ‘However, we can argue that the open ocean convection in the Southern Ocean is actually hindering the simulation of a realistic water masses distribution.’ – This should be shown to be true also for the PI simulation. If it is not, the authors need to argue for why it is reasonable to include it in the LGM when it is not necessary or an improvement to do so for the PI.

We agree with the reviewer that the intense open ocean convection simulated in the Southern Ocean should be detrimental for both time periods, but as the deep water formation via open ocean convection also seems to depend on the background climate (see Fig. 8), we could argue that simulating the right convection processes may seemingly be more critical to get a realistic Atlantic streamfunction at the LGM than it is at the PI. The relatively small differences in the streamfunction obtained with a “PI brines” run (see Fig. S5 and our response to the fourth comment of reviewer #1) illustrate this. After all, models are tuned at the PI, which is also why it may be possible for them to simulate reasonable bottom water properties via the wrong process – which basically amounts to an compensation of errors. As Heuzé (in review, 2020) explains: “In CMIP5 models, no model assessed by Heuzé et al. (2013) could represent dense shelf overflows correctly. Consequently, models relied on open ocean deep convection for their deep water formation. The right amount of deep convection in the Weddell Sea was required for accurate bottom properties; models that convected too little or too much were the most biased.”

Still, as pointed out by reviewer #1 (and now underlined in the text, see our response to the third comment), the parameterization of brines is idealized, so we do not want to define it as default (i.e. standardly used for the PI as well). There is little open ocean convection in simulations with it since the transfer of salt to the bottom grid cells allows for dense water formation without advection being explicitly computed. This sensitivity test is useful to show that open ocean convection might be detrimental to a realistic representation of both the AMOC and the sea-ice seasonality. Nonetheless, it does not solve the issue of our model convecting too easily instead of forming dense shelf water overflows.

R2.E5 : P. 11, L. 349-350: 'showed that few progresses have been made by some modelling groups with respect to that aspect.' - I do not quite understand this sentence. Do you want to say "a few modelling groups have made some (minor?) progress in this aspect", or that "few modelling groups have made any progress in this aspect", or maybe that "some modelling groups have made particularly little progress in this aspect"?

This sentence is indeed phrased in a confusing way. We meant that Heuzé (in review, 2020) showed that some significant progress has been made by a few modelling groups, but as these modelling groups are a minority (among CMIP6 participating groups), simulating the correct convection process at the origin of deep-water formation remains an issue. We have therefore made the following modifications:

As underlined by Heuzé et al. (2013), models struggle to simulate the correct bottom water properties even in the present-day conditions, as they tend to form AABW by open ocean convection, a process rarely observed, instead of the overflow of dense continental shelf water. While none of the CMIP5 models were able to simulate the latter, Heuzé (in review, 2020) showed that a few CMIP6 models are now able to simulate AABW formation via shelf processes, notably thanks to the development of an overflow parameterization. Despite this progress, the issue remains, as "the large majority of climate models form deep water via open ocean deep convection, too deep, too often, over too large an area" (Heuzé, in review, 2020).

R2.E6 : P. 13, L. 409-411: 'It would therefore seem that the correct simulation of convection processes is paramount, and far more important than the choices of boundary conditions, such as the ice-sheet reconstruction [. . .]' – The brines parameterization has not been tested with the PMIP2 boundary conditions, as far as I can tell. Hence, it is clear that it is more important than the choice of ice-sheet reconstruction, but I am not sure it is well founded to say that it is more important than the choice of boundary conditions in general.

Please refer to our response to the last point of reviewer #1.

Figures

R2.F1 : P. 6, L. 153: Somewhat confusing that Fig. 6 is mentioned before Fig. 2. General comment (e.g. figures 3, 4, 7, 8, S2, S3): How are the basins defined (longitudinal and latitudinal limits)? The latitudinal limit for the Southern Ocean seems to be different in different figures (see Fig. 7).

- P6, L153: This sentence is actually not very informative at this place of the manuscript. We have removed it and instead added a reminder at the beginning of the model-data comparison of the sea-ice edges:

Figure 6 presents the simulated sea-ice edges alongside the sea-ice contours based on marine core data, using the reconstruction method described in Sect. 2.4. The sea-ice edge – set at 15% of sea-ice concentration by convention (NSIDC) – of all LGM simulations shows a roughly circular regional distribution around Antarctica (also see Fig. S4).

- The ocean basins are defined using a mask (see Fig. 2 below).

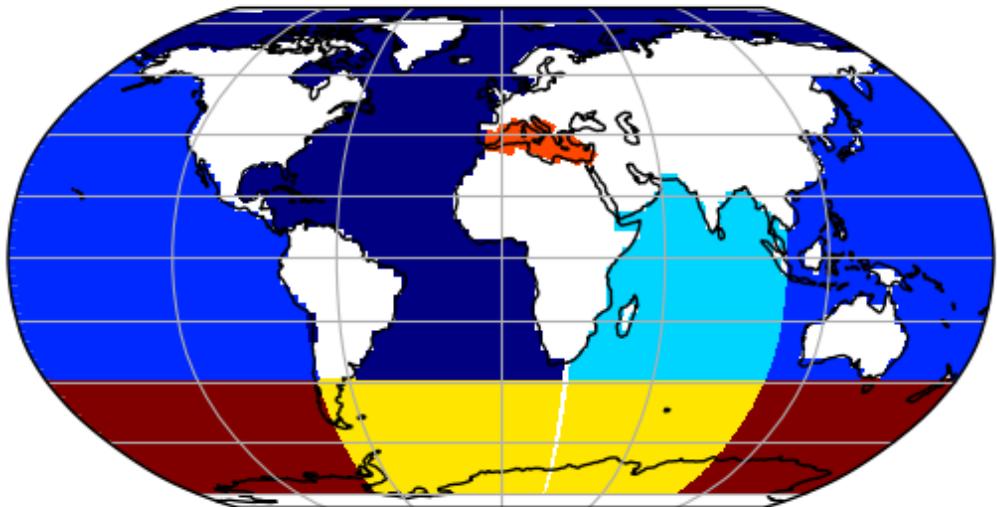


Figure 2: Mask used to defined ocean basins

- For all figures with the exception of Fig. 7 (namely for Fig. 3, 4, 8, S2, S5), the Southern Ocean is set at 36°S. We wanted to get this limit relatively closer to the tip of Africa and of Australia than to the Cap Horn latitude in order not to “cut in half” the interesting warm bias observed in our analysis in the Atlantic and Indian sector. However, it is true that this causes a small discrepancy with the latitudinal limit of 32°S in Fig. 7. The latter cannot be changed easily as the streamfunction computation is done in the code during the run. It should also be noted that it is impossible to use the exact same latitudinal limit to compute the streamfunction and averaged temperatures since the streamfunction is computed using variables on the velocity grid, while the temperature outputs are on the tracer grid. Indeed, the grid of the CLIO model is a B-grid type of the classification of Arakawa (Goosse et al., 2010).

R2.F2 : Figure 2: Please specify in the caption what the mean SST of this simulation is, as this is specified for all other simulations in the figure. Also, it would be advisable to add a few longitude and latitude grid ticks, at least in the bottom row and left column respectively, since you have drawn the grid lines.

We have added in the caption this forgotten value, thank you for pointing that out. The longitudes and latitudes grid ticks will be added to the final version of the figures for the manuscript.

R2.F3 : Figure 4, Panels a-b: In the Methods section 2.5, you say that PI simulations are compared to WOA data, not MARGO data, Caption: Describe the thinner dashed lines surrounding the 1:1 line and how the SSTs are averaged in this figure.

We have corrected this mistake in the x-axis on Fig.4a,b (and also in Fig.S4). Moreover, the caption now reads:

Figure 4. Austral summer (JFM) and winter (JAS) sea-surface temperatures of the Southern Hemisphere in a model versus data diagram, for a sample of our simulations (see Fig. S2 for the complete results). The simulated SSTs are plotted against the SST data from the regridded product (MARGO Project Members (2009) or World Ocean Atlas (1998)) thanks to the aggregation of the coordinates on the nearest ocean grid cell. The 1:1 line features a perfect model-data agreement (black dashed line), while the grey dotted lines features a 5°C departure from it. The marker style indicates the ocean basin of each core. The marker color shows the latitude of the core, except it is

white where the model simulates sea ice in the Southern Ocean. The uncertainties associated with the SST data are plotted by the grey horizontal bars.

R2.F4 : Figure 6, General: I would suggest adding a few longitude and latitude grid ticks, Panel a: Why is the red line in panel a dashed, when no other lines in the figure are?, Caption, L. 1: Please mention how the sea-ice edge relates to the extent and area, Caption, L. 2: The part about the arbitrary index is a bit difficult to read. I suggest "[...] as an arbitrary index on a blue to white scale, where blue denotes no indication of sea ice in proxies, and white denotes agreement of several proxies on the presence of sea ice."

The longitudes and latitudes grid ticks will be added to the final version of the figures for the manuscript. We used a dashed red line in Fig. 6a because we wanted to draw the reader's attention to the fact that the summer contour (drawn based on scarce data) is less well-constrained than the reconstructed winter sea-ice line in Fig. 6b. This is now explained in the caption. Moreover, we thank the reviewer for the suggested clarification of the index explanation. A reminder on how the sea-ice edge relates to the sea-ice extent (shown in Figure 5) has also been added. The sea-ice edge is not related to the sea-ice area (see definition at the beginning of Sect. 3.3). The caption now reads:

Figure 6. Austral summer (a) and winter (b) sea-ice edges (at 15% of sea-ice concentration, enclosing the total ocean surface defined as the sea-ice extent) in the Southern Ocean for the LGM simulations. The sea-ice presence suggested by marine cores data is represented as an arbitrary index on a blue to white scale, where blue denotes no indication of sea ice in proxies, and white denotes agreement of several proxies on the presence of sea ice. The red lines mark the likely delimitation of the sea-ice presence according to the proxy data (compilation of data from Gersonde et al. (2005), Allen et al. (2011), Ferry et al. (2015), Benz et al. (2016), Xiao et al. (2016), Nair et al. (2019), and Ghadi et al. (2020)). We used a solid red line in (b) but a dashed line in (a) as the summer contour is not well-constrained (see Sect. 2.4).

R2.F5 : Figure 7: This figure seems to have a different limit for the Southern Ocean compared to other figures (see General comment). I found this colour scale not very gentle on the eyes. I had trouble looking at the figure because the stark contrast and particularly the bright cyan/mint made me feel dizzy/nauseous. Changing it is of course not a requirement for publication, just a suggestion.

Please refer to our response to your first comment on the figures for the reasons behind the different latitudinal limits. As for the color scale, a more classic red/blue one is now used.

R2.F6 : Figure 8: In my opinion, it would be preferable to plot these using a standardized grid spacing for each column, as it would make it easier to compare the slopes for the different cells

We have standardized the grid spacing and limits to improve readability. To be able to do that, we inverted the y-axis of the two anticlockwise cells. This is now mentioned in the caption.

Minor details (typos and similar)

R2.G1 : P. 1, L. 8: 'inaccurate' – replace for 'inaccurately'

R2.G2 : P. 1, L. 10: 'data-model' – 'model-data' seems to be the more commonly used term, and is also what you use later in the paper

R2.G3 : P. 1, L. 15: ‘water masses properties’ – replace for ‘water mass properties’. Note! This error occurs throughout the manuscript when water mass properties/distributions are mentioned (see e.g. P. 2, L. 24: ‘water masses distribution’ – should be ‘water mass distribution’, as on P. 1, L. 1)

R2.G4 : P. 2, L. 29: ‘shallowing’ – replace for ‘shoaling’

R2.G5 : P. 11, L. 337: ‘paleodata’ – replace for ‘paleoproxy’

R2.G6 : P. 13, L. 399 and 400: ‘sea ice and SST data’, ‘proxy data’ – replace ‘data’ by ‘reconstructions’

We have corrected these typos as advised. Thank you for noticing them.

Supplementary information

R2.H1 : P. 1, first sentence: ‘remplaces. . .’ – should be ‘replaces’

R2.H2 :P. 1, fourth bullet point: ‘all the critical traits stay open’ – ‘traits’ should be ‘straits’

We have corrected these two typos.

R2.H3 :Fig. S1: I would suggest adding a few latitude and longitude grid ticks

The longitudes and latitudes grid ticks will be added to the final version of the figures for the manuscript.

R2.H4 :Fig. S2: See comments for Fig. 4

Please refer to our response to your comments for Fig. 4.

References

Abe-Ouchi, A., Saito, F., Kageyama, M., Braconnot, P., Harrison, S. P., Lambeck, K., Otto-Bliesner, B. L., Peltier, W. R., Tarasov, L., Peterschmitt, J.-Y., and Takahashi, K.: Ice-sheet configuration in the CMIP5/PMIP3 Last Glacial Maximum experiments, *Geosci. Model Dev.*, 8, 3621–3637, <https://doi.org/10.5194/gmd-8-3621-2015>, 2015.

Annan, J. D. and Hargreaves, J. C.: A new global reconstruction of temperature changes at the Last Glacial Maximum, *Clim. Past*, 9, 367–376, <https://doi.org/10.5194/cp-9-367-2013>, 2013.

Antonov, J. I., Seidov, D., Boyer, T. P., Locarnini, R. A., Mishonov, A. V., Garcia, H. E., Baranova, O. K., Zweng, M. M., and Johnson, D. R.: World Ocean Atlas 2009, Volume 2: Salinity, Levitus, S., NOAA Atlas NESDIS 69, US Government Printing Office, Washington, D.C., 2010.

Argus, D. F., Peltier, W. R., Drummond, R., and Moore, A. W.: The Antarctica component of postglacial rebound model ICE-6G-C (VM5a) based on GPS positioning, exposure age dating of ice thicknesses, and relative sea level histories, *Geophys. J. Int.*, 198, 537–563, <https://doi.org/10.1093/gji/ggu140>, 2014.

Benz, V., Esper, O., Gersonde, R., Lamy, F., and Tiedemann, R.: Last Glacial Maximum sea surface temperature and sea-ice extent in the Pacific sector of the Southern Ocean, *Quat. Sci. Rev.*, 146, 216–237, <https://doi.org/10.1016/j.quascirev.2016.06.006>, 2016.

Bé, A. W. H. and Hutson, W. H.: Ecology of planktonic foraminifera and biogeographic patterns of life and fossil assemblages in the Indian Ocean, *Micropaleontology*, 23, 369–414, <https://doi.org/10.2307/1485406>, 1977.

Böhm, E., Lippold, J., Gutjahr, M., Frank, M., Blaser, P., Antz, B., Fohlmeister, J., Frank, N., Andersen, M. B., and Deininger, M.: Strong and deep Atlantic meridional overturning circulation during the last glacial cycle, *Nature*, 517, 73–76, <https://doi.org/10.1038/nature14059>, 2015.

Bouttes, N., Paillard, D., and Roche, D. M.: Impact of brine-induced stratification on the glacial carbon cycle, *Clim. Past*, 6, 575–589, <https://doi.org/10.5194/cp-6-575-2010>, 2010.

Bouttes, N., Roche, D. M., Mariotti, V., and Bopp, L.: Including an ocean carbon cycle model into iLOVECLIM (v1.0), *Geosci. Model Dev.*, 8, 1563–1576, <https://doi.org/10.5194/gmd-8-1563-2015>, 2015.

Cortese, G. and Prebble, J.: A radiolarian-based modern analogue dataset for palaeoenvironmental reconstructions in the southwest Pacific, *Marine Micropaleontology*, 118, 34–49, <https://doi.org/10.1016/j.marmicro.2015.05.002>, 2015.

Crosta, X., Sturm, A., Armand, L., and Pichon, J.-J.: Late Quaternary sea ice history in the Indian sector of the Southern Ocean as recorded by diatom assemblages, *Mar. Micropaleontol.*, 50, 209–223, [https://doi.org/10.1016/S0377-8398\(03\)00072-0](https://doi.org/10.1016/S0377-8398(03)00072-0), 2004.

Esper, O. and Gersonde, R.: New tools for the reconstruction of Pleistocene Antarctic sea ice, *Palaeogeography, Palaeoclimatology, Palaeoecology*, 399, 260–283, <https://doi.org/10.1016/j.palaeo.2014.01.019>, 2014.

Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., Bianchi, C., Bingham, R. G., Blankenship, D. D., Casassa, G., Catania, G., Callens, D., Conway, H., Cook, A. J., Corr, H. F. J., Damaske, D., Damm, V., Ferraccioli, F., Forsberg, R., Fujita, S., Gim, Y., Gogineni, P., Griggs, J. A., Hindmarsh, R. C. A., Holmlund, P., Holt, J. W., Jacobel, R. W., Jenkins, A., Jokat, W., Jordan, T., King, E. C., Kohler, J., Krabill, W., Riger-Kusk, M., Langley, K. A., Leitchenkov, G., Leuschen, C., Luyendyk, B. P., Matsuoka, K., Mouginot, J., Nitsche, F. O., Nogi, Y., Nost, O. A., Popov, S. V., Rignot, E., Rippin, D. M., Rivera, A., Roberts, J., Ross, N., Siegert, M. J., Smith, A. M., Steinhage, D., Studinger, M., Sun, B., Tinto, B. K., Welch, B. C., Wilson, D., Young, D. A., Xiangbin, C., and Zirizzotti, A.: Bedmap2: improved ice bed, surface and thickness datasets for Antarctica, *Cryosphere*, 7, 375–393, <https://doi.org/10.5194/tc-7-375-2013>, 2013.

Galbraith, E. and de Lavergne, C.: Response of a comprehensive climate model to a broad range of external forcings: relevance for deep ocean ventilation and the development of late Cenozoic ice ages, *Clim. Dynam.*, 52, 653–679, <https://doi.org/10.1007/s00382-018-4157-8>, 2019.

Gersonde, R. and Zielinski, U.: The reconstruction of late Quaternary Antarctic sea-ice distribution —the use of diatoms as a proxy for sea-ice, *Palaeogeography, Palaeoclimatology, Palaeoecology*, 162, 263–286, [https://doi.org/10.1016/S0031-0182\(00\)00131-0](https://doi.org/10.1016/S0031-0182(00)00131-0), 2000.

Gersonde, R., Crosta, X., Abelmann, A., and Armand, L.: Sea-surface temperature and sea ice distribution of the Southern Ocean at the EPILOG Last Glacial Maximum — a circum-Antarctic view based on siliceous microfossil records, *Quat. Sci. Rev.*, 24, 869–896, <https://doi.org/10.1016/j.quascirev.2004.07.015>, 2005.

Goosse, H., Brovkin, V., Fichefet, T., Haarsma, R., Huybrechts, P., Jongma, J., Mouchet, A., Selten, F., Barriat, P. Y., Campin, J. M., Deleersnijder, E., Driesschaert, E., Goelzer, H., Janssens, I., Loutre, M. F., Morales Maqueda, M. A., Opsteegh, T., Mathieu, P. P., Munhoven, G., Pettersson, E. J., Renssen, H., Roche, D. M., Schaeffer, M., Tartinville, B., Timmermann, A., and Weber, S. L.: Description of the Earth system model of intermediate complexity LOVECLIM version 1.2, *Geosci. Model Dev.*, 3, 603–633, <https://doi.org/10.5194/gmd-3-603-2010>, 2010.

Goosse, H., Roche, D. M., Mairesse, A., and Berger, M.: Modelling past sea ice changes, *Quat. Sci. Rev.*, 79, 191–206, <https://doi.org/10.1016/j.quascirev.2013.03.011>, 2013.

Heuzé, C., Heywood, K. J., Stevens, D. P., and Ridley, J. K.: Southern Ocean bottom water characteristics in CMIP5 models, *Geophys. Res. Lett.*, 40, 1409–1414, <https://doi.org/10.1002/grl.50287>, 2013.

Heuzé, C.: Antarctic Bottom Water and North Atlantic Deep Water in CMIP6 models, *Ocean Sci.*, <https://doi.org/10.5194/os-2020-66>, in review, 2020.

Howard, W. R. and Prell, W. L.: Late Quaternary surface circulation of the Southern Indian Ocean and its relations to orbital variations, *Paleoceanography*, 7, 79–117, <https://doi.org/10.1029/91PA02994>, 1992.

Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., and Ames, D. P.: Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption, *Environmental Modelling and Software*, 119, 32–48, <https://doi.org/10.1016/j.envsoft.2019.05.001>, 2019.

Kageyama, M., Albani, S., Braconnot, P., Harrison, S. P., Hopcroft, P. O., Ivanovic, R. F., Lambert, F., Marti, O., Peltier, W. R., Peterschmitt, J.-Y., Roche, D. M., Tarasov, L., Zhang, X., Brady, E. C., Haywood, A. M., LeGrande, A. N., Lunt, D. J., Mahowald, N. M., Mikolajewicz, U., Nisancioglu, K. H., Otto-Bliesner, B. L., Renssen, H., Tomas, R. A., Zhang, Q., Abe-Ouchi, A., Bartlein, P. J., Cao, J., Li, Q., Lohmann, G., Ohgaito, R., Shi, X., Volodin, E., Yoshida, K., Zhang, X., and Zheng, W.: The PMIP4 contribution to CMIP6 – Part 4: Scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments, *Geosci. Model Dev.*, 10, 4035–4055, <https://doi.org/10.5194/gmd-10-4035-2017>, 2017.

Kageyama, M., Harrison, S. P., Kapsch, M.-L., Löfverström, M., Lora, J. M., Mikolajewicz, U., Sherriff-Tadano, S., Vadsaria, T., Abe-Ouchi, A., Bouttes, N., Chandan, D., LeGrande, A. N., Lhardy, F., Lohmann, G., Morozova, P. A., Ohgaito, R., Peltier, W. R., Quiquet, A., Roche, D. M., Shi, X., Schmittner, A., Tierney, J. E., and Volodin, E.: The PMIP4-CMIP6 Last Glacial Maximum experiments: preliminary results and comparison with the PMIP3-CMIP5 simulations, *Clim. Past*, <https://doi.org/10.5194/cp-2019-169>, in review, 2020.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., Zweng, M. M., and Johnson, D. R.: World Ocean Atlas 2009, Volume 1: Temperature, Levitus, S., NOAA Atlas NESDIS 68, US Government Printing Office, Washington, D.C., 2010.

Lynch-Stieglitz, J., Adkins, J. F., Curry, W. B., Dokken, T., Hall, I. R., Herguera, J. C., Hirschi, J. J.-M., Ivanova, E. V., Kissel, C., Marchal, O., Marchitto, T. M., McCave, I. N., McManus, J. F., Mulitza, S., Ninnemann, U., Peeters, F., Yu, E.-F., and Zahn, R.: Atlantic Meridional Overturning Circulation During the Last Glacial Maximum, *Science*, 316, 66–69, <https://doi.org/10.1126/science.1137127>, 2007.

MARGO Project Members: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, *Nat. Geosci.*, 2, 127–132, <https://doi.org/10.1038/NGEO411>, 2009.

Orsi, A. H., Johnson, G. C., and Bullister, J. L.: Circulation, mixing, and production of Antarctic Bottom Water, *Prog. Oceanogr.*, 43, 55–109, [https://doi.org/10.1016/S0079-6611\(99\)00004-X](https://doi.org/10.1016/S0079-6611(99)00004-X), 1999.

Peltier, W. R.: Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE, *Annu. Rev. Earth Pl. Sc.*, 32, 111–149, <https://doi.org/10.1146/annurev.earth.32.082503.144359>, 2004.

Peltier, W. R., Argus, D. F., and Drummond, R.: Space geodesy constrains ice age terminal deglaciation: The global ICE-6G-C (VM5a) model, *J. Geophys. Res.-Sol. Ea.*, 120, 450–487, <https://doi.org/10.1002/2014JB011176>, 2015.

Prahl, F., Herbert, T. D., Brassell, S., Ohkouchi, N., Pagani, M., Rosell-Melé, A., Repeta, D., and Sikes, E.: Status of alkenone paleothermometer calibration: Report from Working Group 3, *Geochem. Geophys. Geosyst.*, 1, <https://doi.org/10.1029/2000GC000058>, 2000.

Roche, D. M., Dokken, T. M., Goosse, H., Renssen, H., and Weber, S. L.: Climate of the Last Glacial Maximum: sensitivity studies and model-data comparison with the LOVECLIM coupled model, *Clim. Past*, 3, 205–224, <https://doi.org/10.5194/cpd-2-1105-2006>, 2007.

Roche, D. M., Crosta, X., and Renssen, H.: Evaluating Southern Ocean sea-ice for the Last Glacial Maximum and pre-industrial climates: PMIP-2 models and data evidence, *Quat. Sci. Rev.*, 56, 99–106, <https://doi.org/10.1016/j.quascirev.2012.09.020>, 2012.

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthélemy, A., Benshila, R., Chanut, J., Levy, C., Masson, S., and Vivier, F.: The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities, *Geosci. Model Dev.*, 8, 2991–3005, <https://doi.org/10.5194/gmd-8-2991-2015>, 2015.

Shin, S.-I., Liu, Z., Otto-Bliesner, B. L., Kutzbach, J. E., and Vavrus, S. J.: Southern Ocean sea-ice control of the glacial North Atlantic thermohaline circulation, *Geophys. Res. Lett.*, 30, <https://doi.org/10.1029/2002GL015513>, 2003.

Sikes, E. L., Volkman, J. K., Robertson, L. G., and Pichon, J.-J.: Alkenones and alkenes in surface waters and sediments of the Southern Ocean: Implications for paleotemperature estimation in polar regions, *Geochimica et Cosmochimica Acta*, 61, 1495–1505, [https://doi.org/10.1016/S0016-7037\(97\)00017-3](https://doi.org/10.1016/S0016-7037(97)00017-3), 1997.

Tarasov, L., Dyke, A. S., Neal, R. M., and Peltier, W. R.: A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling, *Earth Planet. Sc. Lett.*, 315–316, 30–40, <https://doi.org/10.1016/j.epsl.2011.09.010>, 2012.

Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., and Poulsen, C. J.: Glacial cooling and climate sensitivity revisited, *Nature*, 584, 569–573, <https://doi.org/10.1038/s41586-020-2617-x>, 2020.

US National Snow and Ice Data Center: NSIDC, <https://nsidc.org/>.

Watterson, I. G.: Non-dimensional measures of climate model performance, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 16, 37–391, [https://doi.org/10.1002/\(SICI\)1097-0088\(199604\)16:4<379::AID-JOC18>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-0088(199604)16:4<379::AID-JOC18>3.0.CO;2-U), 1996.

Williams, G. D., Aoki, S., Jacobs, S. S., Rintoul, S. R., Tamura, T., and Bindoff, N. L.: Antarctic Bottom Water from the Adélie and George V Land coast, East Antarctica (140–149E), *J. Geophys. Res.*, 115, <https://doi.org/10.1029/2009JC005812>, 2010.

World Ocean Atlas: <http://ingrid.ldeo.columbia.edu/SOURCES/.NOAA/.NODC/.WOA98/>, 1998.