We thank both reviewers for thoughtful/constructive responses.

**Response to review/comments by John Andrews**

*But for many readers the important point that this paper make is in section 4.4 "Brief comparison to past geological inferences" is indeed brief a mere 8 lines but this statement outlines the other important, indeed critical, verification of the modelled ice sheets and their expansion and retraction, that is the aerial extent of the ice sheets, a necessary but not sufficient parameter in the calculations of ice volume an global sea level. It is a call for action to the glacial geological community, however, the problems have not changed significantly since the Clark et al 1993 paper—that is the ability to provide a date on buried stratigraphic units, primarily tills, that are older than the 50,000 radiocarbon dating limit—this problem remains.*

Yup, very brief. The intent, as noted, is to prompt the glacial geology community with some model-based chronologies. Some of the reviewer's comments will in part be incorporated/addressed into the revised text but we will also explicitly refer the reader to this review for the valuable discussion and to ensure appropriate academic credit.

*so it is difficult to see why the growth and decay should be more symmetrical.*

Whether the saw-tooth paradigm is appropriate for shorter stadial/interstadial transitions is unclear to us. Yes a large LGM NA ice complex with extensive warm based regions should have fast and strong deglacial intervals, but whether this is necessary for MIS5d is unclear. The model used in this study, does lack shallow-shelf approximation ice physics (now addressed in ongoing work), so grounding line retreat is poorly represented. But the terrestrial components should have more confidence. The revised submission will give more guidance on model uncertainties and how they should be taken into account by readers.

*Some figures are too small.*

Will address in revision.

**Response to review by Andre Ganopolski**

1. Phase space of last glacial inception

*The meaning of the term "phase space of last glacial inception" which authors put in the title and mentioned several times in the text, is unclear to me. Since "phase space" is space, their dimensions (axis) should be properly defined. For example, for mechanical systems, phase space is defined by coordinates and momentum. For the climate system, Fig 3b in Ganopolski et al. (2016) gives an example of another phase space. Here the position of glacial inceptions (bifurcation point) is shown in the insolation–CO2 phase space. The authors should either clearly define what they understand under "phase space" in their manuscript or abandon this term. A similar situation is with the term "bifurcation" which authors used several times (p. 16 and 22) but the meaning of this term remains unclear.*

Fair enough for some sloppy useage. We clearly aren't using the standard physics definition for phase space. We are debating between replacing with "trajectory space": or defining phase space as the "4D space of possible ice/climate histories over glacial inception as represented by LCice". Bifurcation has been used in the popular sense of division into disjoint branches with respect to some characteristic (and not in the technical sense of dynamics systems theory). We'll add a footnote clarifying this.

2. Introduction

*The authors devoted less than one page for discussing previous modelling works related to the last glacial inception. Apart from several own papers, they only cited my publications (Calov et al. (2005); Calov et al. (2009 ) and Ganopolski et al. ( 2010)) and the only information Bahadory et al. provide about our works is the spatial resolution of the CLIMBER-2 model: "The model used in that study employed very low resolution (51longitude by 10 latitude for atmosphere and approximately 100 km for the ice sheet model)" (page 2, line 47).*

For the record, the claim is incorrect. Our submission also states "the one coupled ice/climate modelling study that adequately captured both the growth and retreat phases of LGI required the use of an imposed (albeit plausible) aeolian dust deposition forcing and temperature bias correction (Ganopolski et al., 2010)"

*"On other hand, it should be noted that the relative quality of modelled LGM ice extent in Ganopolski et al. (2010) attests the potential value of using fast EMICS like CLIMBER"*

*Amazingly, just 20 lines below (page 3, line 66) the authors again describe the spatial resolution of CLIMBER-2: "A further limitation in this latter study is that the CLIMBER EMIC employed uses a 2.5D statistical-dynamical atmospheric model with very limited longitudinal resolution (51.4o) and a 3 basin 2D ocean model". Do the authors realize that they describe the resolution of the same model twice?*

Oops, sorry about that. Now fixed. (Submission was rushed to make the IPCC deadline, based on North American time...)

*In any case, I believe that the manuscript by Bahadory et al. will benefit a lot if, apart from the resolution of CLIMBER-2, the authors will discuss also other relevant publications. Already in Calov et al. (2005), we cited in the introduction more than 25 modeling papers and since then the number of relevant publications increased significantly.*

We agree that the submission would benefit from more review of past work, and will do so (especially for more recent papers). However, we also note that some of the "25 modeling papers" cited in Calov et al. (2005) used what we judge to be poor model/experimental configuration/designs and obtained poor results in large discord with paleo proxy constraints. Some also just used flow-line models and/or otherwise lacked 2D geographic resolution. For these cases, we see no point in referencing.

3. Temperature biases and realistic simulations of ice sheets extent

*The authors stress in the manuscript that they do not use any climate bias corrections and I fully agree that bias correction represents a trade-off between internal consistency and the realism of past climate simulations. All climate models have biases and for simulations of quasi-linear response of the system (like CO2 increase), climate biases are likely not very important. However, for simulation of glacial inception, which is a fundamentally nonlinear process, temperature biases can be much more important because their magnitude can be comparable with the climate response to orbital forcing. In their previous paper (Bahadory and Tarasov, 2018), the authors wrote a lot about temperature biases but provided no information about spatial patterns and magnitude of temperature biases. Table 3 only indicates that average temperature over the box covering most of Canada is close to reality. The real problem is, however, not average but strong (5 to 10oC depending on the season) zonal temperature gradient over northern North America related to the atmospheric circulation and explained by quasi-stationary planetary waves. Due to the coarse spatial resolution of CLIMBER- 2, this effect is not resolved and this leads to a strong, dipole-like temperature bias (Ganopolski et al., 2010; Fig 2a). This is why it is noteworthy that the North American ice sheets simulated without temperature bias correction in Calov et al. (2005a) (Fig. 6) and in Bahadory et al. (Fig. 4 and 5) are very similar with the thickest ice located over Alaska. Introducing of temperature bias correction in Ganopolski et al. (2010) led to a very different ice sheet evolution which we believe is more realistic. The similarity between Bahadory et al. and our old results (Calov et al. , 2005) can be caused by the fact that the LOVECLIM model, in spite of a higher spatial resolution, has a rather simplistic atmospheric model which results in similar to CLIMBER-2 temperature biases. At least, this is what one can conclude from Fig. 1b in Heinemann et al. (2014), another paper based on the LOVECLIM model. By the way, in this paper temperature bias correction has been used. Bias correction has been used also in a number of GCMs studies such as Vizcaino et al. (2008); Herrington and Poulsen (2012). This is why, it would be useful to show present-day (preindustrial) summer temperature biases simulated by LOVECLIM model used by Bahadory et al. This can be, for example, the average value over the 50 ensemble members or a single representative one. Of course, it is up to the authors to decide which technique to employ but they should inform their readers about potential serious drawbacks.*

Good point. And as seen below, the NA present-day biases are large for Loveclim over the range of models that passed our acceptance threshold for glacial inception. However, even with the large present-day warm bias, Baffin is one of the first places to glaciate in our model.

Eurasian biases are relatively much less (not shown, but will be in the revised submission). We will be adding a much more complete discussion on model limitations and how this should affect interpretation of results. This will include plots of subensemble means and variances of present-day bias.

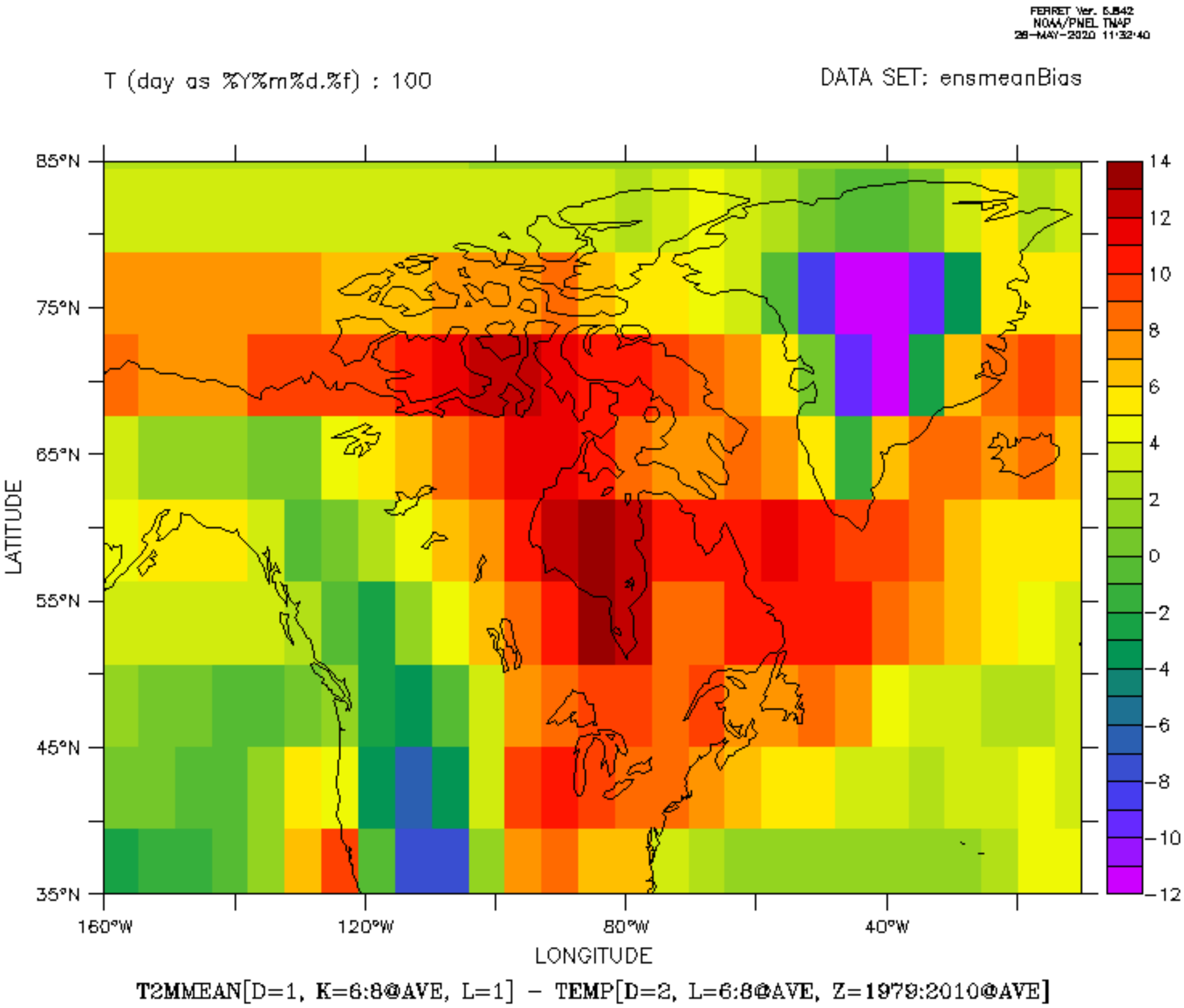


Figure: Passed subensemble JJA mean surface air temperature bias.

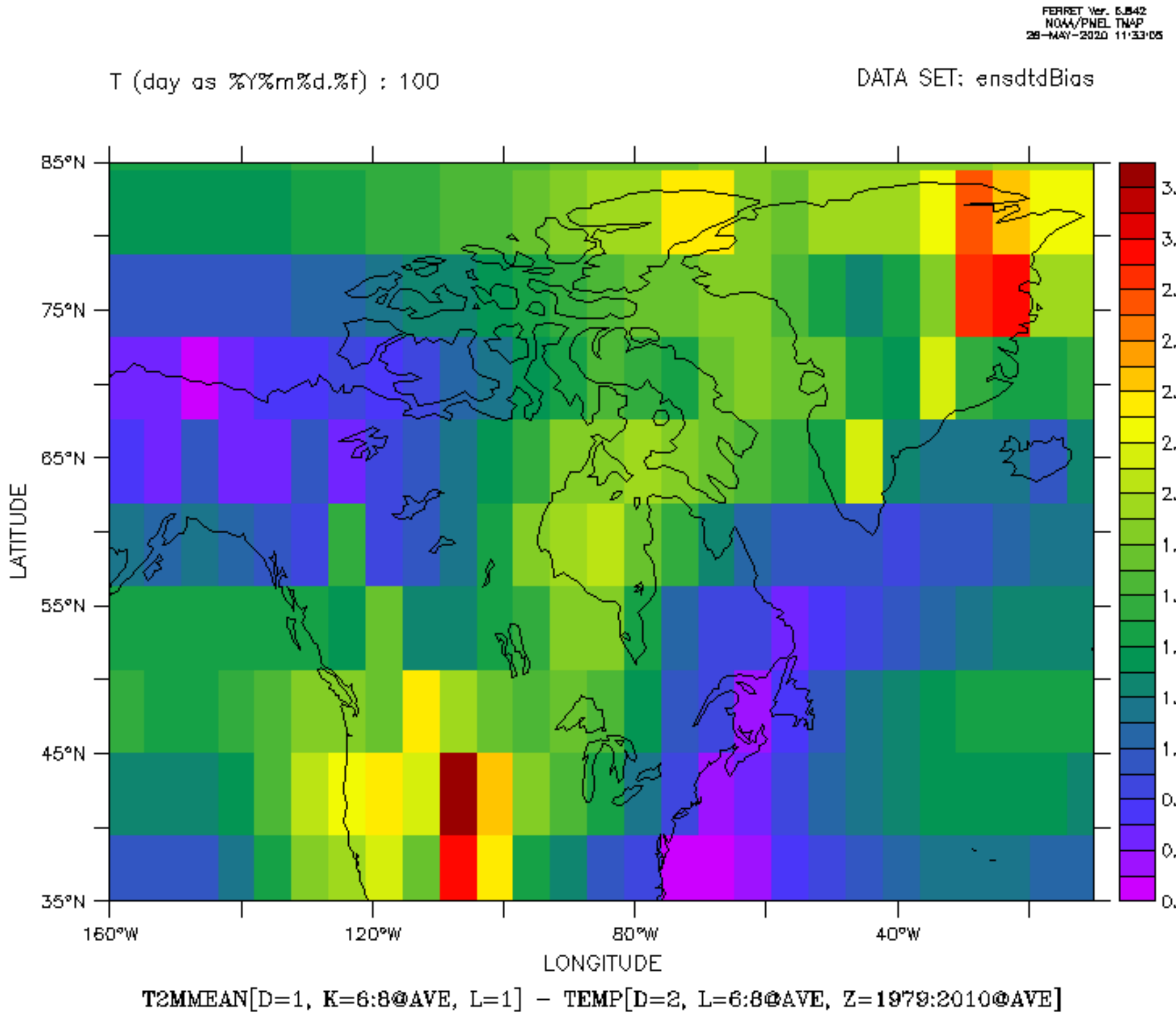


Figure: passed subensemble standard deviation JJA surface air temperature

4. Present-day constraints on model parameters

*On page 10 the authors wrote that "despite having different start times (due to different calendar start years between 122 ka and 119 ka ...), all simulations start growing ice in the first 100 years of simulation". It is not clear from the paper which runs started at which time, as well as why start time was chosen differently for different runs. However, the fact that according to Fig. 1 the model simulates between 10 to 20 meters sea level drop already at 119 ka is worrisome. Indeed, since climate before 120 ka was similar to preindustrial one or even warmer, such rapid ice sheet growth at the beginning of model runs indicates that at least some model realizations would simulate glacial inception already during the late Holocene which, of course, is in odd with observational data. In Section 5, Bahadory and Tarasov (2018) wrote that they used "a trial criteria based on ice volume changes (between 1700 and 1980 CE)" to reject model versions which grow "too much" ice during this* *interval. But this interval is much too short for such a test. For example, Fig.3 in Bahadory et al. clearly shows how much ice is formed after year 200 since the beginning of the runs. Since summer insolation and GGHs concentrations remained practically constant at least since 1000 BCE till ca. 1900 CE, testing of whether or not selected model versions simulate glacial inception in the late Holocene would require at least 10 times longer runs than have been performed by the authors. To be able to judge their realism, it is crucially important to know how much "present-day" ice is simulated by different model versions.*

We respectivefully disagree on the present-day test interval being to short. The interval was appropriate for the given context of extracting an ensemble of 500 model runs closest to equilibrium mass-balance out of 2000 model runs.

As to the question of whether the models are positive mass-balance biased, that is already clearly the case for most models from figure 10 of Bahadory and Tarasov (2018). The more fundamental question of to what extent this distorts the results of the present work is in good part answered by the result of having models that subsequently retreat post-stadial at a rate that is consistent with sealevel proxies (within relevant proxy uncertainties).

The Northern mid to high latitude Eemian summer insolation maximum occured around 126 ka and Lisieki/Raymo 2004 have their Eemian sealevel highstand (likely dominated by Antarctica) at 123 ka. So it may well be that North America started growing ice earlier than the 120-122 ka our model runs started at. As such, the late start time likely offsets some (or all?) of the impact of the present-day positive surface mass-balance bias.

The reviewer also raises the point that start times for the passed subensemble don't have their actual start times listed. We will add a list of the subensemble parameter vectors to the revised supplement.

5. Spatial patterns of simulated North American ice sheet

*When discussing spatial patterns of simulated Laurentide ice sheet, the authors wrote "to our knowledge, there is no community-based geologically-inferred MIS 5 ice margin reconstruction for NA. Aside from the issue of Alaska (and certain adjacent parts of the Yukon), our results are, within (large) age uncertainties, consistent with the till stratigraphy presented in Clark et al. (1993)" (page 22). However, Clark et al (1993) explicitly stated that "the Laurentide Ice Sheet first developed during Stage 5 over Keewatin, Quebec and Baffin Island" (page 79) which is inconsistent with the results presented by the authors. It is also noteworthy that the recent reconstruction of NH ice sheet for the MIS 5d presented in Batchelor et al. (2018), also places MIS 5d Laurentide ice sheet over northern-eastern Canada and implies very little glaciation over Alaska and in Western Canada. Since I am not an expert in the history of glaciation of North America during the last glacial inception, I wonder what the authors think about these apparent inconsistencies?*

*And if they really believe that there are no reliable reconstructions for the ice sheets during MIS 5d (I do not understand the meaning of "community-based"), then what is the motivation for performing a large ensemble of transient last glacial inception simulations?*

Clark et al. (1993) do not discuss Ellesmere and they differentiate Laurentide from the Cordilleran ice sheet. As for the Western Canadian Arctic, their discussion is summarized in fig 16, which indicates no constraints during MIS5d:c. Furthermore, their summary figure 16 shows no MIS5d glaciation over Quebec, contradicting the statement in the abstract. So the above quote from Clark et al, in combination with their figure 16 is consistent with the results in our fig 3 and fig 5. We will add some of these details to the discussion in the revision.

We would also argue that Batchelor et al. (2018) best estimate MIS5e in their figure 1 is inconsistent with sealevel constraints. Their use of a single scaling estimate for ice volume is not appropriate when you have multiple domes nor an ice sheet that is unlikely to be in equilibrium. 3 circular ice caps of the same total area as a single circular ice cap will have less total ice volume. Discrepancies can get even larger when you use have non-circular geometries. We therefore find this reconstruction problematic, but will look more carefully at their cited constraints, to see what aspects are more robust and address this in the revisions.

6. Using simulations of glacial inceptions to constrain transient climate response

*The authors devoted only one paragraph in the text to the description of how they used their model ensemble to constrain transient climate response (TCR). However, they highlighted this result in the abstract where they suggested that their results can be used "to constrain future climate change". Since future climate change is a very hot issue, this small part of the manuscript deserves serious attention, especially, because the authors put a very tight constraint on TCR (0.7-1.4 C). If their estimate of TCR is correct, then only five of ca. 30 different GCMs participating in CMIP5 have the right TCR while all other overestimate it. Obviously, such a claim has very serious implications for future climate change projections.*

The reviewer should have quoted the full sentence from the abstract: "This therefore underlines the potential value of fully coupled ice/climate modelling of last glacial inception to constrain future climate change". IE, we do not claim that our modelling results should be used to constrain climate sensitivity only that there is potential value to do so from fully coupled ice/climate modelling.

*Below I argue why simulations of glacial inception cannot constrain future climate change. Although a number of attempts have been made to use paleoclimate data and results of paleoclimate modelling to constrain equilibrium climate sensitivity (ECS), these studies cannot directly constrain TCR. Indeed, although there is some correlation between TCR and ECS, this correlation is not very tight and TCR of* *different models with similar ECS can differ by factor two. The reason is that TCR strongly depends on the rate of ocean heat uptake which differs significantly between climate models. Obviously, simulations of glacial inception provide no constraints on ocean heat uptake. This is why below I only discuss whether simulations of glacial inception can constrain ECS.*

*i) Climate sensitivity to CO2 doubling (ECS) and the response of climate to seasonal and latitudinal redistribution of insolation are caused by completely different forcings an, therefore, numerous processes and feedbacks play a completely different role. I am not aware of any study about the relationship between regional and seasonal climate response to insolation change and global climate response to CO2 change (ECS), but I doubt whether there is a strong correlation between these very different climate changes.*

*ii) As far as the simulated rate of ice sheet growth is concerned, the situation is even more complex because ice sheet growth is controlled not only by simulated climate change but by many other factors. The first one is model biases in modern climatology. If these biases are comparable with climate response to orbital forcing, then there is a big question of whether ice sheets growth can constrain future climate change. Second, ice sheet response to orbital forcing strongly depends on surface mass balance parameterization. The authors used the PDD scheme which does not even explicitly account for the direct impact of orbital forcing on the surface mass balance of ice sheets and a number of studies (e.g. van de Berg et al., 2011; Bauer and Ganopolski, 2017) questioned the applicability of this simplistic scheme to the modelling of ice sheet response to orbital forcing.*

There is no "the PDD scheme", and the scheme we use (temperature dependent melt coefficients derived from energy balance modelling) is different than what most have used to date and arguably indirectly does take into account SW dependencies better than the common PDD scheme with fixed degree-day melt coefficients. It still though does not have explicit dependence on solar insolation and we will make this caveat explicit in the revised text. It should also be noted, that a surface mass-balance scheme with explicit surface insolation dependence has been implemented in the GSM in 2019 (and is now the default, but came too late for the ensembles in this project).

*In short, I do not believe that simulations of glacial inception can really constrain ECS, let alone TCR*

This is a fair critique (and what should be be obvious in hindsight learning on TCR versus ECS) and we have now computed the ensemble ECS for comparison (and will replace TCR with ECS examination in the revisions). As per the figure below, the requirement of capture of last glacial inception and subsequent retreat still provides some constraint on ECS, rejecting runs with ECS < 1.3 C. But now this criteria does not provide an upper bound constraint, in contrast to that for TCR. However, this figure also shows the limited range of ECS probed over the current ensemble (which us being addressed in ongoing work). ECS will depend on the radiative forcing of 2*CO2 (which varies somewhat across models) as well on internal feedbacks. The reviewer fairly points out that response to orbital changes in insolation will be subject to different feedbacks than that for future 2*CO2. However, some of these feedbacks will be similar (eg snow and sea-ice albedo). Furthermore, last glacial inception also included changes in pCO2.

There is also a submission (Choudhury, Simulating Marine Isotope Stage 7 with a coupled climate-ice sheet model) that recently completed TCD discussion of relevance. It used LoveClim radiative forcing dependence on CO2 as one of its two ensemble parameters. Again, capture of the stadial/interstadial response strongly narrowed down parameter ranges. For this case, I would therefore expect a strong correlation between model ECS and stadial/interstadial capture.

ECS test of full ensemble (0ka to 3ka). Simulations using parameter vectors that passed glacial inception constraints are in black.