Leger et al. present simulations of the Greenland Ice Sheet's history spanning the last 24,000 years. The simulations rely on a newer generation ice sheet model run at 5 km resolution and forced by a climatology from a recent transient simulation of the last deglaciation (ICESM: 21-11ka), and a climate index scheme using snapshot periods to extend this climate history back to 24ka and from the early Holocene to the year 1850 CE. An ensemble of simulations is presented, with variations in key model parameters that influence ice flow, surface mass balance, ocean-ice interaction, and climate. The results of these simulations were compared against a recent gridded reconstruction constraining the timing of GrIS retreat across the last 24kyr, which is constrained by spatiotemporally varying available geologic proxies (PaleoGris).

Generally, I found the manuscript to be well written, and in most places, easy to follow. I do consider the model-data comparison framework presented here a hallmark of this work. Prior ice sheet modeling studies across similar intervals have either compared results to point measurements or regional reconstructions past ice sheet retreat (margin reconstruction), relative sea-level records, or no comparison at all. This work constitutes to my knowledge the first attempt to compare simulated ice sheet modeling results to a complete reconstruction of deglacial ice margin change (of course acknowledging the uncertainties in PaleGrIS as well). By doing so the authors constrain the state of the GrIS at a few key intervals adding knowledge as to the possible GrIS areal extent and volume during the LGM (or local LGM), drivers of deglacial retreat when many portions of the GrIS resided on or near the continental shelf, and GrIS Holocene evolution up to 1850CE. The authors do a good job acknowledging shortcomings in their modeling approach, highlighting that no unique simulation exists that satisfies goodness of fit with the geologic reconstructions during all time intervals. Nevertheless, the results and model outputs are an important contribution and should have wide appeal. While I am supportive of publication, I do have some points below that should be addressed before publication.

General Comments:

- From my understanding of the text, it seems the simulations were found to be most sensitive to climate (temperature and precipitation and their offsets) and the ice flow enhancement factor. Other parameters seemed to have a reduced influence on the simulated ice state and retreat. As a reader, I took this to mean that climate biases and simulated ice rheology may produce large misfits in the simulated ice history. While these uncertainties are generally expressed well in the text, few things stick out to me that could be better acknowledged in the text

(*Note, please also see line by line comments for specific places where the text could be more clear).

- From my knowledge working with collaborators that use ICESM and older versions like Trace21ka, there are some potential warm biases in ICESM temperature during many intervals such as the LGM and Younger Dryas (https://ui.adsabs.harvard.edu/abs/2023AGUFMPP14B..01T/abstract). This bias was shown in Badgeley et al., (2020; https://doi.org/10.5194/cp-16-1325-2020). I point the authors to Figure 13 for an example of Trace21ka temperature vs. reanalysis.
- In paleo ice flow modelling, the climate forcing is one of the most uncertain pieces necessary to simulate ice history over paleoclimate timescales. I would recommend the authos plot their derived temperature forcing (such as that shown in Figure 5) against the Badgeley et al. (2020) product which is available here: https://arcticdata.io/catalog/view/doi%3A10.18739%2FA2599Z26M. This may provide additional information as to if the +/- 3.5 degrees Celsius magnitude temperature offset used in the ensemble was enough to capture climate uncertainty.
- *1850 CE ice extent*: I was a bit surprised to see the 1850 CE ice extent in Northern Greenland be so extensive, although given the difficulties in simulating paleo ice history this is not a major criticism. Instead, I am curious if the simulated 1850 CE state being too extensive outside of present-day ice margin is a consequence of the large magnitude cooling following peak temperatures in the climate forcing (Figure 5). Ice core proxies and reanalysis suggest a lower magnitude of cooling following the HTM, which may limit regrowth from Holocene minimum.

- Simulated basal ice temperature: I did find it odd that the ice sheet remains temperate, for the exception of the margins throughout the entirety of the transient simulation. While PISM may have a warm bias from looking at MacGregor et al. (2016;2022), I wonder if biases in the simulated ice sheet temperature are contributing to the ensemble derived high sensitivity to ice flow enhancement. If the simulated ice sheet is too warm, that may require the low enhancement values (to make ice stiffer), which the authors showed were necessary to have a better match to PaleoGris during certain time intervals. If this is a possibility, it would be nice to see this better acknowledged in the text. See below in the line-by-line comments for specific location to clarify this if my assessment is logical.

Specific Comments:

Line 71: Why not cite Lecavalier et al. (2014) here? I believe this reconstruction is used for the GIA reconstruction in IMBIE assessments currently (HUY3 model) and is to some degree data constrained (RSL records). I believe these other simulations listed do not do any appreciable model-data comparison.

Line 88: change "…with a paleoclimate, and ii)…" to "…with a paleoclimate reconstruction, and ii)…"

Line 215 (Figure 2): In the "Red Text" line near the bottom of figure, "(results/discussion)" is highlighted. There does not seem to be a defined results or discussion section in this paper. Maybe instead, highlight the specific sections here (e.g. 3.1-4.1)

Figure 4: I leave it up to the authors here, but I would find temperature anomalies from a reference period (e.g. 1850 CE) to be more informative about the degree of spatial variability.

Line 483: What is reference height temperature? Is that the 2-meter temperature at the height of the ice sheet in the climate model output?

Line 488: "…we use data…". Replace 'data' to 'output'

Line 492-501: I am confused with the climate forcing setup when looking at the text and Figure 5. If possible, perhaps this paragraph could be rewritten so that the climate forcing timeseries is written sequentially from older to younger time periods? For example, "….between 24ka-21ka we use…, between 21ka-11ka we use transient output from ICESM, between 11-9ka we use…., and 9ka to 1850CE we use."

Line 505: The authors use sea-surface temperature as the oceanic forcing. This is not ideal as research supports that shelf depth temperature is more influential for ice-ocean interaction in Greenland. Likewise, the authors could have used simulated shelf depth temperature from ICESM, similar to Tabone et al., 2024 (https://www.nature.com/articles/s41467-024-50772-5), since they used modeled outputs of salinity bridged by linear interpolation in this modelling. While I recognize that there is a lack of palaeoceanographic reconstructions at depth, especially over such timeframe, this shortcoming should be at least acknowledged here in the text and discussion (which I will highlight further down).

Line 505: What is the reason for starting from 24 kyr BP? Is it possible that staring at 21 kyr BP, when transient output starts for ICESM would have yielded similar results for the simulations?

Line 668: Make sure citations in chronological order (O Cofaigh).

Line 692: Doesn't Figure 9 show the periphery glaciers (Test 3)? The text states periphery glaciers are removed.

Line 715: Can Table 1 be moved up closer to the text here?

Line 846: I do wonder to what extent the poor fit in NE Greenland is due to the use of SST's from Osman et al.? Tabone et al. (2024; nature.com/articles/s41467-024-50772-5) simulate the ice margin in this region using a different ice sheet model, forced by Trace-21ka shelf depth temperature. Would it be possible to compare the ocean temperatures used in this study across this region to shelf depth temperatures from ICESM? Perhaps anomalies would be best to compare against.

Page 24: There are references to a number of geographic locations. If possible, it would be nice if some of these locations could be placed on one of your Greenland maps (abbreviated), for example on Figure 14.

Line 940-950: Very interesting to see the relationship to model resolution. I would also acknowledge that each model used different climate forcings, ice flow approximations. Also, do these models use different model extents for the LGM mask (max. ice extent)?

Line 1194 and 1270: For consistency, would it be better to change title in line 1194 to "….during the last deglaciation" instead of late glacial to better match the subtitle on line 1270?

Line 1205-1210: This result is similar to Tabone et al. (2018; https://cp.copernicus.org/articles/14/455/2018/), so would be useful to cite.

Figure 16: Should the subtitle read "5 best lLGM simulations" instead of "5 best LGM simulations"?

Line 1276: I cannot find the '13-12 kyr BP' time slice in supplemental figure 2.

Line 1373: I think Briner et al., 2014 is the wrong citation here as that was more for west central Greenland. Larsen et al. (2015; https://doi.org/10.1130/G36476.1) and Larsen et al. (2011; https://doi.org/10.1016/j.quascirev.2011.07.022) would be more appropriate citations here.

Line 1377-1382: It would be good to know what parameter combinations were responsible for the ~100 km retreat (high PDD factors?).

Lines 1521-1544: The thinning curves go back to ~11.5 ka. Any reason why the analysis was only for the 8kyr to 1850CE period and not back further since this section does include discussion about the early Holocene? It would have been nice to see if any models simulated the early Holocene thickening.

Lines 1546-1555: It is interesting that the LGM and 1850 CE simulated basal regime has the majority of the ice sheet at pressure melting. I am assuming the margins are colder based because the ice is thinner there, but I would have expected to see some modeled transient behavior of the ice sheet through time. Looking at MacGregor et al., 2016;2022, it does seem compared to other ice sheet models, PISM may have more of a warm bias. However, at the ice divides where vertical advection dominates, it's surprising not to see any cold based ice. Is there any explanation for this?

Lines 1610-1620: I really do wonder if the climate forcing used is really responsible for some of this mismatch. See some of my other comments, but ICESM has a warm bias compared to ice core proxies with it being warmer during the Younger Dryas than proxies show. Additionally, the trend in Holocene temperature is not that similar to proxies (or Data assimilated products like Badgeley et al., 2020; Buizert et al., 2018), which reconstruct the HTM earlier in the Holocene, followed by a slight cooling towards stable temperature, whereas the reconstructed climate used here has peak warming at ~ 6 ka followed by a large magnitude of cooling to 1850CE.

Section 4.3:

- The temperature and precipitation offset were shown to have a large influence. Doesn't this mean that the climate forcing uncertainty plays a major role? It would be interesting to see where the reconstruction for temperature used in this work compares against data assimilated products which are available (Badgeley et al., 2020).
- Additionally, the flow enhancement was shown to be important, requiring 'more rigid' ice to expand during the LGM (line 1635). Could this be a consequence of the simulated ice temperature (at least we see the basal

temps in supplemental figures) being too warm, and therefore needing adjustment in ice flow enhancement?

Line 1652: I agree an ensemble approach is beneficial, but future work could also make use of more climate reconstructions. Here I think the climate uncertainty is under sampled compared to the parameter uncertainty (personal opinion though).

Line 1732: Again, I think ICESM has an LGM warm bias (or at least should be looked into). See Badgeley et al. (2020; Figure 13). At least Trace21ka had a warm bias during the LGM compared to data assimilated products and ice core proxies.

Lines 1785-1790: Downs et al. (2020; https://tc.copernicus.org/articles/14/1121/2020/) used data assimilation techniques, sampling climate and model parameter uncertainty, to determine that higher precipitation than modern was likely needed to simulate Holocene margin migration in SW Greenland. Might be a good citation here to acknowledge that other modeling studies have come to a similar conclusion indicating a need for increased precipitation.

Lines 1795-1805: I am totally fine with the use of the Osman SST records as forcing for the ocean model. But it would be good to acknowledge here that it is not ideal, and shelf depth temperature would be more accurate (even though that does not exist).

Lines 1808-1815: Is there any information on the bias of ICESM precipitation for present day?

Lines 1820-1825: See comment above. Downs et al. (2020) may be a good citation supporting enhancement of precipitation anomalies needed to match geologic reconstructions.

Section 5.2: This paper is long (and for good reason with all of the analysis), so maybe this would not make a big impact. However, since you did not simulate the GrIS with multiple mesh resolutions, perhaps this portion of the text could go in the supplemental, and you can instead shorten this by citing your analysis of bedrock topography at different resolutions (if you move the majority of this text and figure to the supplement)? Also, Cuzzone et al. (2019; https://tc.copernicus.org/articles/13/879/2019/) highlighted the influence of mesh resolution on paleo ice retreat (albeit in 1 region of Greenland) and ice mass change. I would think 5km should be good enough, but in areas of very complex bed topo, or where gradient in the surface mass balance is important to simulate, higher resolution would have a larger impact on mass flux and comparison to ice margin migration.

Line 2019: If you plot your reconstruction against data assimilated products (Badgeley et al., 2020; Buizert et al., 2018), you might find that the +/- 3.5 degrees Celsius offset is not enough to account for uncertainty in temperature.