Reply to the Comments from Referee #1 (Report #2)

We thank Referee #1 for the time and effort he/she spent reading the manuscript again and providing further suggestions for improvements. Please find below a point-by-point response to the referee's comments which are shown in italics.

SIE vs SIA

Presumably you've calculated SIE as the total area of all grid cells with at least 15% SIC. However, "sea ice extent is a strongly grid-dependent, nonlinear quantity" (Notz et al., 2020). That paper also points out that the observational spread for trends in SIA is smaller than it is for trends in SIE, so strongly advocates the use of SIA over SIE in this sort of investigation. I think it's critical to state whether ORAS5 (your comparison data set) has the same grid as your HN model output, because if it doesn't then your HIRHAM-NAOSIM SIE values are not comparable to the ORAS5 ones (even though they may be close). Just looking at Zuo et al, it looks like the ECV data coming out of ORAS5 might be given on a 1x1 degree grid whereas yours is much finer? In light of that, and particularly as your grid cells are non-uniform in size, it would be much better for this paper to present its results in terms of SIA. I don't think it will change your results much (although it might reconcile the HIRHAM/NAOSIM trends to ORAS5 a bit?. But it will make the results more rigorous and adhere to SIMIP best practice in terms of model output intercomparison. This change should be made prior to publication.

[REPLY] We have followed the referee's suggestion and have replaced SIE by SIA throughout the manuscript. In addition, we now mention in the figure captions that "ORAS5 data were bi-linearly interpolated to the NAOSIM grid using distance-weighted averages for missing values".

Subjective Language

The word "significant" appears eleven times in this manuscript. It's used twice as part of the phrase "statistically significant" (which is of course fine), but in the other nine times I think it's used quite subjectively and should be reconsidered. Many journals now (in my view rightly) don't allow this. For example, Nature journals now do not allow "significant" to be used without an accompanying p-value. I've already mentioned that describing your changes as a "major improvement" also seems subjective - I think it's safe to say things have improved, but whether your improvements are major vs minor is perhaps in the eye of the beholder. The same with the description of the "simple, but innovative cloud-cover dependence". I agree it's simple, but whether it's really that innovative or not is probably something for the cover-letter rather than peer-reviewed literature (so I suggest you remove that word).

[REPLY] We have removed words like "significant" or "significantly" when we are not explicitly referring to a statistical test (three times). In the other case, we have added the attribute "statistically" in order to clearly indicate that the significance was statistically tested (seven times). Furthermore, we have removed the attributes "major" for the improvement and "innovative" for the cloud-cover dependence.

ORAS5 biases

You should discuss some implications of using ORAS5 as a benchmark from which to calculate bias. Per your rebuttal, you're implicitly treating it as "the truth" for purposes of measuring whether HNnew is better than HNold. I think doing that is fine, but it needs context.

[REPLY] We have expanded the introduction of ORAS5 as reference data set for the evaluation of sea-ice volume and sea-ice area (lines 199–208 in the revised version). We refer now to biases in ORAS5 as discussed by Tietsche et al. (2018) and Zuo et al. (2019), and we explicitly point to differences in sea-ice volume between ORAS5 and PIOMAS.

For example regarding Figure 6, there is a report which claims ORAS5 is too thick in winter. How do biases in the product affect your findings regarding the model? What's the risk that the albedo params are being tuned to match a dataset that is itself biased?

[REPLY] Indeed, ORAS5 tends to overestimate the thickness of sea ice as indicated by Tietsche et al. (2018); on the other hand, PIOMAS rather tends to underestimate the ice thickness as indicated by Schweiger et al. (2011). Assuming that we had used PIOMAS as benchmark, the bias reduction in HNnew compared to HNold would have been even more obvious, considering that HNold shows even thicker sea ice than ORAS5. This means that there is no risk when we claim that HNnew performs better than HNold.

Could you also elaborate on what you mean on line 211? I understand that model biases are larger than the spread of observational products, but I don't understand what you mean about "qualitatively equal model biases would appear" if you evaluated against the observational products.

[REPLY] This statement has been completely removed from the manuscript.

Initial Conditions

I'd still like to know more about the "initial conditions" i.e. is the snow realistically deep, is the ice in a realistic place and does it have realistic thickness. From L100 I gather that P1 was initialised from a long-term run of NAOSIM but the others from ORAS5. For P1 I'm less concerned about the initial conditions because of the spinup time. But I think it is relevant for P2 & P3: was the initial snow depth distribution also taken from ORAS5? For P2 & P3 it would be good to see a supplemental two-panel figure with a map of (a) initial SIT distribution (b) initial snow depth distribution. I imagine the SIT distribution is sufficiently realistic, but I do think it's important when experimenting on such a short timescale to know about the initial snow depth. I.e. how deep is it and where is it. Because that may affect the timing of snow melt onset, which is one subject of the paper.

[REPLY] In the first round of the peer review process, it was requested to demonstrate that the findings do not depend on the specific initial conditions. Therefore, an alternative method of initializing the model was chosen for P2, namely using (more realistic) ORAS5 fields instead of (more consistent) model restart fields as in P1. In contrast to P1, P2 is initialized with zero snow thickness. We explicitly mention this fact now in line 92.

As snow rarely survives the warm season, the snow thickness distribution in all simulations, be it from P1 or from P2, is similar after the first melting season due to the applied nudging to ERA5, which leads to similar snow fall patterns in the model. Whether the simulated snow distribution is realistic or unrealistic is an open issue. Ongoing studies show that the evaluation of the modeled snow fall or snow distribution represents a fundamental problem, not only from the model side but also from the observational side. Tackling this problem goes far beyond the scope of this manuscript.

In contrast to snow, sea ice possesses some kind of memory effect, namely in terms of multi-year sea ice. The initial ice thickness distribution is therefore not irrelevant for a few years. The fact is, the initial ice thickness distribution is completely different in P1 and P2, and nonetheless, the effect of the revised snow albedo parameterization in both pairs is comparable. We have added maps of the initial ice thickness distribution in P1 and P2 as a supplemental figure to the manuscript. Showing the initial snow thickness distribution is redundant due to the aforementioned reasons, and showing any initial conditons from P3 is redundant as well, because the 10-year-long spinup time is long enough to forget any initial state.

Other

L218: it would be good to get some numbers in here for slopes. At what rates are HNnew, HNold & ORAS5 declining?

[REPLY] Corresponding numbers have been added (lines 214–215 in the revised version).

L220: I guess this is because of the nudging scheme? I.e. the place where sea ice is, and is not, is driven by atmospheric circulation. But the thickness is driven by the radiative balance.

[REPLY] Only P1 and P2 were carried out with nudging, because these simulations were primarily intended for the comparsion with MOSAiC measurements. The P3 simulations (which are discussed in this sections) were running without nudging, meaning that the atmospheric circulation not only drives the sea ice, but also responds to the sea-ice conditions. Nevertheless, we agree that the ice area is more strongly controlled by the atmospheric circulation than the ice volume.