The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2019-314-RC1, 2020

Interactive comment on "An inter-comparison of the mass budget of the Arctic sea ice in CMIP6 models" *by* Ann Keen et al.

Anonymous Referee #1

Received and published: 27 March 2020

General Comments

This paper is a first look at the Arctic ice mass budget terms saved by participating CMIP-6 modeling groups. The seasonal cycle for these mass budget terms during a reference period (1990-2009) is examined and differences between models discussed. Historical simulations concatenated by SPS5-85 projections are analyzed to examine differences in the evolution of total extent/mass and budget terms from 1960 through the end of the 21st century. The core finding is that, despite substantially different seasonal cycles and time trajectories, the differences in individual mass budget terms between models are relatively small. The paper is well written and constitutes a substantial effort in coordination between many different groups to provide output that can be easily compared. This effort is laudable and the results should provide a useful

reference as presented. As a reviewer, my job is to check the paper on originality, scientific quality, impact, and presentation quality. I think the paper is good to excellent in these categories but could gain some more impact through more focused analysis and presentation. As an author I am typically annoyed when reviewers ask me to write a paper I didn't mean to write so please take the following as suggestions. While a useful compilation of results as presented, the paper tends to scratch only the surface on some questions and falls a bit short of a deeper exploration. For example, a discussion of how the relatively small differences in the budget terms lead to rather large different mean states and trajectories would be interesting. This might perhaps be achieved by looking at cumulative budget differences (or anomalies from ensemble mean) so that one could see the relative contribution of budget terms to the state trajectory? Some of the inter-model differences seem to be obscured by the way they are presented. For example, some inter-model differences could perhaps be highlighted if shown as the change between periods (reference to late 21st century) instead of showing the full transition (Fig 4, 11). Maybe highlighting individual models as extremes might be useful to tease out some fundamental differences between models (or groups of models)? For example, the NCAR-CESM2 and CanESM-5 seem to have rather dramatic differences in their summer and winter trajectories, with the NCAR-CESM2 showing a very early decline in the summer and CanESM-5 being an outlier in its winter trajectory. This seems to be in part related to the very strong basal melt in CanESM-5? Figure 11, looking at mean state vs. flux terms, seems to indicate some clustering of models that may allow drawing conclusions about differences between models or groups. Some of the graphics could probably be condensed to highlight differences and focus on a particular thesis. The "forced" model run section could use some work to better relate it to the coupled runs and allow a sharpening of the conclusion beyond " both model sea-ice physics and atmospheric forcings are important". If this is too difficult maybe it should be cut entirely.

I hope you will find the below suggestions useful and recognize that they may not necessarily work or be too difficult to implement (or even make sense).

Specific Comments:

Abstract: Maybe a note about the lack of fundamental diversity in model physics would be useful here as a caveat. Line 81: Keen and Blockley (2018). Is it worth to add a section on whether or not any of the results presented here are different or the same as in your previous paper? Maybe a sentence in the discussion or conclusion? Line 110: Can you explain why "basal growth" and "basal melt" aren't the same quantity just with opposite sign? Line 121:"SSP5-8.5 scenario" . . . given the debate about the likelihood of that scenario it might be useful to caveat that using an extreme scenario may actually be useful to highlight differences in the budget terms. Line 137: Why are both HadISST versions used? Is this supposed to provide some measure of the uncertainty in the ice concentration data? Uncertainty for PIOMAS might also be quantified in this context. Line 163.. "at least one of the observational data set". . . see above comment about uncertainty in the observations. Line 176: "All models have their seasonal maximum ice mass in May" consistent with PIOMAS". . . This must be due to the choice of domain since the full PIOMAS domain has its maximum volume in April not in May. Some note, either here or earlier on about how the domain can affect some of the variables though not likely the general conclusions, should be made. Figure 3: I think it would be better to show observations with some uncertainty measure (relates to above). PIOMAS uncertainty could be scaled from values given by Schweiger et al. 2011 to the domain? Why is the summer total ice mass trajectory not shown in 3b? Can you comment on the large differences in the summer/winter trajectories for NCAR-CESM2 vs. CanESM5? NCAR-CESM2 seems to be losing its ice much more rapidly in summer while CanESM-5 loses its winter ice much more rapidly than the other models? Line 199: "whereas other models show a more uniform decline". Don't you think this is likely because of the 2100 cut off and if you'd looked at beyond 2100 you'd see a similar flattening towards the end? Line 208. Figure 4b. Maybe it would be more informative to plot delta sea ice/delta temperature (sea ice sensitivity) between two periods. That would highlight model differences better. Line 211: "1960-1986" Maybe a statement why the 1960 through 1986 reference" period is chosen is different from the period over which Figure

2 seasonal cycle values are averaged (1990-2009). I suppose you were looking for greatest overlap with observations for the Figure 2 comparison. Line 230, figure 3b.. "20%" there is no percentage scale given that this number could be related to. Can you add a percentage axis in 3b or rephrase to one doesn't go looking for it. Fig 6,7,8 I find it pretty difficult to see the differences between models as plotted and the information density pretty low. Plotting seasonal cycles/trajectories relative to a multi-model mean would probably highlight differences better. Absolute values as shown in Figure 6, if deemed relevant, might be put into a table? I am looking for some visuals that highlight the conclusions drawn regarding each of the subsequent paragraphs. The reader has to do quite a bit of visual hunting to relate the statements made to the graphics. Line 292. . . "there is considerable variability between the models" again, I think removing the multi-model mean and showing this as anomalies relative to that mean would highlight the differences.

Line 309: "how the ice state impacts the evolution of each budget term". As said in the general comments, a discussion of how the budget terms accumulate to the evolution of the state would be useful. Line 314.. move (Fig 9a) to just after "area of ice" , maybe add a references (Bitz and Roe 2004) to link to think ice growth feedback process. Line 317: What does "in-situ" mean in this context? That sounds a bit off. Figure 9b: Heading says "Surface Melt" elsewhere named "Top Melt"? Fig 9/10 b/c. A little bit more discussion why the trajectories for surface and basal melt go in opposite direction for per unit area vs. total (9 bc/10bc) would be helpful. Something that states that while the melt at the base and surface of the remaining ice increases the total area of ice decreases, so there is less ice to which this process applies. I know you say this in Line 318 but maybe this could be made a bit clearer. Pointing to Figure 10, and maybe discussing the extreme case of CanESM5 would make this clear right away. I was scratching my head for a bit until this sunk in (maybe me). Fig 11. Wondering if this figure wouldn't be more informative if not every point of the trajectory were plotted for each model but rather the change from reference period to some target period (end of 21st) or before. This would probably highlight the model differences better. Line 385

Mass budgets for forced runs. Wondering if this shouldn't be a separate section in the body of the paper rather than in the discussion. I am struggling a bit what to make of this set of experiments and what the fundamental results are. I understand the goal of trying to quantify the relative impact of ice physics vs. forcing on the mass budget. I think the experiments with the same forcing but different physics accomplishes this w.r.t to one a couple of parameters. To make this meaningful though, I think this limited sample of sensitivities would somehow need be related to the spectrum of possible parameters in the CMIP-6 models. Does it represent an extreme and therefore establish a bracket of sea ice physics sensitivities? Similarly, the forcing sensitivities need to somehow be related to the range of atmospheric variables. What is the range between CORE II, DFS5.2 and MetUM coupled (whatever that is?). Of course that's not easy thing to quantify because the forcing consists of multiple variables and there could be compensating differences, but maybe temperature and downwelling radiation would capture a sufficient measure of the range. Of course this would have to be done for the CMIP-6 models then and that could get messy because of the coupling. I don't know what the solution is but I think as presented, this set of experiment probably confuses more than it adds to the discussion.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-314, 2020.