

Dear Referees,

Thank you both for taking the time to read our manuscript and for your useful and constructive comments and suggestions. We are pleased with your positive reviews and have incorporated many of your suggestions, which have helped to improve the manuscript.

Our responses to your comments are included below in blue, together with details of how we have incorporated these into the revised manuscript. References to page and figure numbers refer to the revised manuscript.

Yours sincerely, on behalf of all the co-authors,
Ann Keen.

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?

Thank you for these suggestions. We have updated the discussion of how the model budgets change during the 21st century, and now include plots showing the evolution of the anomalies in the budget terms, both for the multi-model mean (new Fig. 10b) and the

individual models (new Fig. 13). Hopefully these plots show more clearly the relative changes in the different budget terms, and how the balance of these alters as the model integrations progress. This is also discussed in the updated text (For example P16, L639-643)

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).

Fig 4 (Fig 3 in the revised manuscript) is now updated to show only the first panel, and numbers quantifying the changes for two periods relative to 1960-89 are given in the text. We have retained the evolution over the entire period (relative to 1960-89) in the figure as it illustrates the divergence in response between the different models as the 21st century progresses.

For Fig 11 (Fig 14 in the revised manuscript), the aim of this figure is to highlight the underlying similarity in the evolution of the model responses, which is lost if two time periods are chosen as the model budgets change at different rates. We have updated the text in this section, so that hopefully this is more clearly explained (P16, L651-653).

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?

Thank you for this suggestion. We were not able to use the CanESM5 model for this purpose, as some of the main budget terms were unavailable for this model (the frazil and advection components). However, in section 6 we do now highlight models with different ice state trajectories as we discuss the differences between the models. (For example, P15, L584-590)

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.

While we have not specifically used figure 11 for this, we have been able to identify differences between the model budgets during the reference period that are consistent with some of the sea ice parametrization choices. This is explored in the new section 5.

Some of the graphics could probably be condensed to highlight differences and focus on a particular thesis.

We agree and have updated a number of the figures to better illustrate the points made in the manuscript. For example, figure 7 (figure 6 in the updated manuscript) now has a pane for each budget term rather than for each model, to better illustrate the differences between the models. Figures 8,9 and 10 (replaced by new figures 10 and 12 in the updated manuscript) now show decadal means rather than annual means, to show the evolution and differences between the models more clearly. There are further details in the specific responses below.

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.

We have now included a new section ‘Understanding differences between the CMIP6 models’ (Section 5), where the forced experiments are introduced and used to help understand differences between the model mass budgets during the reference period.

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.

This has now been added (P2 L52)

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?

We have now added some text in the discussion to mention that these results are broadly consistent with the findings of Keen and Blockley (2018). (P18, L713-719).

Line 110: Can you explain why “basal growth” and “basal melt” aren’t the same quantity just with opposite sign?

These two terms are defined separately in the SIMIP mass budget and are important at different times of the year, hence we feel it is important to keep them separate in the analysis. The factors affecting each can also be different.

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.

We agree and have added some text to mention this. (P4, L129-133)

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.

Yes, that’s right. We have clarified this in the text and added a comment about PIOMAS uncertainty. (P4, L141-142 and L155)

Line 163.. “at least one of the observational data set” . . . see above comment about uncertainty in the observations.

This sentence has been re-worded (P5, L171-172).

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.

This has been noted in section 2, where the domain is first mentioned (P3, L113-116).

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?

An uncertainty range for the observations has been added to each plot.

We did not include both summer and winter mass trajectories in Fig 3b (Fig. 2b in the revised manuscript) as the summer trajectories did not add any significant extra information: relative differences in the rate of mass decline between the different models are very similar whether looking at the March or September values. We now mention this in the text (P5, L203-4). In addition, the figure would have become quite cluttered as there is not a clear separation between the summer and winter values.

We have added text to highlight the fact that models with the fastest decline in summer ice area are not necessarily those with the fastest winter decline, although we have not identified a particular model feature that may cause this (P5, L200-201).

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?

Yes, we agree and have updated the text accordingly (P6, L211-212).

Line 208. Figure 4b. Maybe it would be more informative to plot delta seaice/delta temperature (sea ice sensitivity) between two periods. That would highlight model differences better.

Thank you for this suggestion. We did generate a plot of the ice sensitivity, which we planned to include. However, as the sensitivity of the CMIP6 models is now discussed in the recently published SIMIP community paper we decided to refer to this instead and have updated the appropriate text accordingly (P6, L225-232).

(SIMIP Community (2020). Arctic sea ice in CMIP6. *Geophysical Research Letters*, 47, e2019GL086749. <https://doi.org/10.1029/2019GL086749>)

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.

Yes, that’s right. For the budget reference period we chose a period during which the ice area and mass are relatively stable. For the seasonal cycles of ice state, we chose a

later period with more observational data (including satellite data). This has been mentioned in the text. (P6, L236-239).

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.

I think this refers to figure 5b (figure 4b in the updated manuscript), and the percentages have now been added to the figure.

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.

Figure 6 (figure 5 in the revised manuscript) has been updated to show only the main budget terms, and to also include the multi-model mean for reference. Figure 7 (figure 6 in the revised manuscript) has been altered so that the seasonal cycles of all the models (for each budget component) are on the same plot. Figure 8 (Figure 13 in the revised manuscript) has been updated to show anomalies relative to the reference period, and decadal means rather than annual means. We feel that these are an improvement over the original versions.

We tried generating versions of both Figure 6 and Figure 7 using anomalies relative to the multi-model mean but found these to be hard to interpret. In addition, one of the aims of figure 6 (figure 5 in the revised manuscript) is to show the broad similarity of the overall balance of the model budget terms, which was lost when plotting the anomalies.

Line292 . . . “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.

The new fig. 6e, while still showing the absolute values, does now show the differences between the different models more clearly as the vertical scale is more appropriate.

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.

We have revised the discussion of the evolution of the ice budget, and the associated figures. In particular, we now show how the anomalies in the main budget terms evolve, both for the multi-model mean and for the individual models, and that different terms are important for the ice loss at different stages in the ice decline. (For example P16, L639-643).

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.

We have added this reference – thank you for mentioning it (P12, L511).

Line317: What does “in-situ” mean in this context? That sounds a bit off. Figure 9b: Heading says “Surface Melt” elsewhere named “Top Melt”?

The figure heading has been corrected, and the text modified. We meant the values expressed ‘per unit area of the ice’, and this has been re-worded in the revised version of this section so that it is clearer. (P13 L514-516)

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).

The discussion of these trajectories been updated and is now illustrated by the multi-model mean (new figure 10). We have improved the text so that hopefully this is clearer, for example the summary to this section contains the following: “... the amount of basal melt per unit area of the ice continues to increase throughout the 21st century (Fig. 10c), but the total amount of ice lost by basal melt reaches a maximum in the 2020s and declines after that because there is a reduced area of ice over which the melting can occur.” (P14, L564-566)

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.

As mentioned above, we have retained Fig 11 (Fig 14 in the updated manuscript), as the aim of this figure is to highlight the underlying similarity in the evolution of the model responses, which is lost if two time periods are chosen as the model budgets change at different rates. We have updated the text in this section, so hopefully this is more clearly explained. (P16, L651-653).

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 to 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.

We have now added a new section to the paper ('Understanding differences between the models'), where the forced runs are introduced and used to help understand the differences in the model budgets. We have added some comments about the sensitivities considered in the forced models, and how these relate to the range of options covered in the CMIP6 models (for example they do not represent an extreme), but as mentioned it is difficult to quantify this. We have been able to identify some common features in the model budgets that are consistent with some of the parametrization choices.

Interactive comment on "An inter-comparison of the mass budget of the Arctic sea ice in CMIP6

models" by Ann Keen et al.

Till Wagner (Referee)

Received and published: 1 June 2020

In this manuscript, the authors compare the individual terms of the Arctic sea ice mass budget (predominantly surface and bottom melt, basal growth and frazil ice formation, and advection) from 14 CMIP6 models. This type of in-depth analysis had not been possible for previous CMIPs, as individual mass budget terms were not routinely reported.

The paper is very well written and structured, clearly illustrated, and the subject matter is a good fit for The Cryosphere. It presents an interesting result in that ~half of the simulated annual ice loss is due to basal melt, and ~one quarter each due to surface melt and advection out of the Arctic. Another central result is that ice formation occurs predominantly as basal growth, with frazil ice formation playing a substantial role depending on a model's minimum frazil ice thickness. Finally, it is striking to see how consistent the partitioning of the individual mass budget terms is between models.

In light of this I recommend the paper for publication after minor revisions.

I agree with the other reviewer with regard to two general comments:

1) I found the paper could do with some more focus, and investigate model differences in more detail. In particular highlighting which differences are due to different physics (e.g., meltpond or radiation schemes) and which are due to different parameter values (such as minimum frazil ice thickness) would be of interest.

This brings me to my main comment. Many of the models share the same sea ice model components (CICE, LIM, SIS) and I believe it would be of interest what the differences are between models with the same sea ice component vs between models with different sea ice components. In the case this is not insightful, it would nevertheless be helpful to discuss the role of having shared sea ice components or not (e.g., for a reader like me who is keenly interested but no expert in the differences between sea ice model components).

2) In line with the other reviewer's comment, I found the section on the forced model runs somewhat vague and only tangentially related to central message of the paper. I would also suggest that this section be either incorporated more carefully or cut (which would further

help focusing the main story).

In response to both your general comments (and those of the other reviewer), we have now included a new section 'Understanding differences between the CMIP6 models' (Section 5), where the forced experiments are introduced and used to help understand differences between the model mass budgets during the reference period. We have also looked more closely at the model budget differences and identified links between some of the budget terms and the formulation of the sea ice component.

Specific Comments:

I.65 "uncertainty" (not uncertainly)

I.68 " models' " (not model's)

I.74 maybe "emerging consensus" rather than "increasing appreciation"

These changes have been made (L70, L74 and L80)

I.135 what time periods are covered by the 3 observational products?

For the seasonal cycles shown in figure 2 (figure 1 in the revised manuscript), we use the years 1990-2009 for each observational dataset.

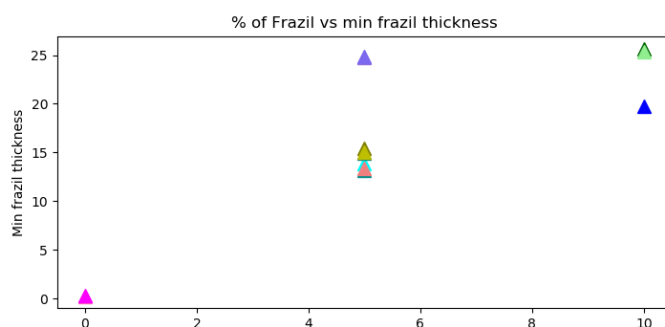
For the trajectories shown in figures 3 and 4 (figures 2 and 3 in the revised manuscript), we now use the following time periods:

- OSI-SAF, HadISST1.2 and HadISST2,2,0,0: 1979-2015
- PIOMAS: 1979-2018
- HadCRUT4: 1960-2019

This is now clarified in the manuscript. (P4, L140-150).

I.263 Is this a linear relationship? Would it be insightful to plot "% of frazil ice formation" vs "min frazil ice thickness"

That is an interesting question. There is some evidence that it may be a linear relationship (see plot below). However, since we submitted this work the term representing frazil ice formation in the MRI-ESM2 model has been added to the basal growth term, which means we no longer have a frazil ice term for a model with a minimum frazil thickness of 20cm. So as there are not many data points we have not focussed on this in the text.



Plot of frazil ice formation (as a percentage of total ice growth) for the reference period 1960 to 89 for each model, plotted against the minimum thickness of frazil ice formation.

I.309 period at end of sentence

I.350 "amount" (not about)

The text in this section has now been updated. The former sentence is now not present, the latter is corrected at L657.

Fig 4 would it make sense to add the observations?

Yes, we have now added observations to Figure 4a. (figure 3 in the revised manuscript).

Figure 4b has been removed, as we instead reference the recently published SIMIP community paper: SIMIP Community (2020). Arctic seaice in CMIP6. Geophysical Research Letters, 47, e2019GL086749. <https://doi.org/10.1029/2019GL086749>

Fig 5b add legend

A legend has been added. (figure 4b in the revised manuscript).

Fig 6 It is difficult to see the differences between the models. Is there a more concise way to present this data? In line with my major comment above, maybe it would be worth grouping the models by sea ice model component? e.g., instead of the bar plots have one subplot for all CICE models, one for all LIM models, etc, with each model's basal growth value indicated by a marker and the same for the other terms?

Fig 7 As for Fig 6 it is difficult to see differences between models.

Figures 6 and 7 have now been updated (figures 5 and 6 in the revised manuscript), and hopefully are now clearer. Thank you for the suggestion to group the models. A new figure 9 shows the main budget terms grouped according to a number of key model parametrizations and settings, and this has proved very useful in helping to understand some of the budget differences.

Fig 9 and 10 are cropped on the left edge.

Fig 9 doesn't show units on vertical axis, Fig 10 does.

Fig 9b is "surface melt", Fig 10b is "top melt".

These figures have been replaced in the updated manuscript. and Fig. 9 is replaced by the new Fig. 10c, which shows the multi-model mean values per unit area of the ice., and Fig 10 is replaced by the new Fig. 12. We have checked to ensure that the units are noted in each case, that the figures are not cropped, and that the headings are correct.

Fig 9a is "basal growth", Fig 10a is "frazil ice formation". Why are the different quantities shown, why not all 4 main terms for each Figure?

Yes, we realise now that this was rather confusing. The figures for this section have been altered, and the evolution of the budget terms is now shown for all the main terms (new Fig. 12). The values per unit area of the ice are now shown for the multi-model mean (new Fig. 10c), and only for those processes that occur over the ice surface. This is explained in the caption.

Fig 11 are the units on the vertical axis Gt year^{-1} rather than kg year^{-1} ?

Yes, thank you for spotting this. It has now been corrected (new Fig. 14)

Fig 12 maybe leave out the lines for lateral melt, snowice, evapsubl, since they are mostly negligible and make the figure harder to read?

Thank you for this suggestion. We have now removed these smaller terms from all the figures apart from figure 4 (where the budget is first introduced). This is a definite improvement.