Response to second review of 'CMIP6 models overestimate melt, growth & conduction fluxes relative to ice mass balance buoy estimates' (Mathieu Plante)

Throughout this response, the original review is reproduced in black; our responses are shown in red. References are given at the end.

Review of: "CMIP6 models overestimate sea ice melt, growth & conduction relative to ice mass balance buoy estimates" by Alex E. West and Edward W. Blockley.

This manuscript discusses the performance of CMIP6 models in reproducing the sea ice thermodynamics as measured from Ice Mass Balance (IMB) buoys in the central Arctic and the Beaufort Sea. In particular, the authors inter-compare the simulated sea ice growth and melt from different CMIP6 members and discuss differences in terms of the simulated sea ice state (yearly mean thickness and area). They effectively group the selected CMIP6 members according to their component characteristics, allowing them to identify common patterns and the source of discrepancies. They find that in general, the simulated thermodynamic fluxes in the selected CMIP6 members respond in a realistic manner to the simulated climatological sea ice state but overestimate the magnitude of these fluxes.

I find that the manuscript is very interesting, pertinent and of good quality. The presented results are insightful on model sensitivities, and on the importance of an accurate representation of the heat fluxes at the air-ice and ice-ocean interfaces. The use of in-situ observations to evaluate the internal sea ice thermodynamics is also a significant contribution. This analysis relevant for publication in the Journal of Model Developments.

Nonetheless, I find that the more general contributions in the manuscript are sometimes difficult to isolate within the detailed results. There are also a few needed corrections and clarifications to better guide the reader towards the main conclusions.

I thus recommend this manuscript to be accepted for publication, after major revisions, as listed below.

We thank the reviewer for their positive and constructive remarks.

Mathieu Plante

General comments:

- There is a tendency to use overly concise wordings, at the cost of preciseness and sometimes accuracy. This is especially present in introduction, where some steps in the reasonings are skipped, likely because they seem obvious to the authors, but effectively leaves it to the reader to work it out. I believe it is worth spending more wordings to be more precisions, especially for processes that are later referred to in the analysis.

We understand the point the reviewer is making. We think there is scope for a fuller discussion of the relationship between ice growth and melt and ice volume within the Introduction, possibly with a schematic, although this may require removal of a figure elsewhere. We expand on this plan below in response to the reviewer's specific requests.

- I would like to have more information in the method sections on the metrics used: e.g., which are diagnostics (i.e., growth and melt rate terms) vs. derived from the internal temperature profiles (conductive fluxes), and how they relate to the method used to get the in-situ values from IMBs. This would also help to shift the focus on the use of IMBs to assess CMIP6 models (for a GMD manuscript).

All assessed metrics are direct model diagnostics, including conductive fluxes, and we will make this clear. This differs from the IMB measurements, where conductive fluxes are derived from temperature measurements – but in this case, melt and growth fluxes are themselves also derived, from elevation measurements.

Of course, some of the individual models themselves will derive conductive fluxes from internal temperatures within their sea ice thermodynamics codes – though many derive full implicit solutions, including temperatures and conductive fluxes, in a single thermodynamic solver.

- I find it difficult to fully discuss ice top heat conduction without discussing the snow layer. This is somewhat covered in section 5, but could be mentioned earlier and throughout the analysis.

In CICE5.1.2 and CICE4, the top heat conduction is defined as the downwards heat flux from the snow/ice surface to the interior of the top layer of the snow-ice column. This top layer may be snow, if snow is present, or ice if snow is not present. When top conductive heat fluxes were derived from the IMB data, we strove to match this definition by measuring the conduction to the surface of the snow-ice column (rather than the top surface of the ice, although these are coincident if no snow is present).

- Discussing the mushy layer model results, the authors vaguely refer to un-reported terms in the basal heat balance. This should be more more specific, otherwise I find it somewhat difficult to interpret these model results. I suspect that the authors are referring to the treatment of sea ice congelation not fully accounting for the conductive heat flux in the CICE mushy layer congelation scheme (see Plante et al., 2024), but I am not sure. If it is so, then it is important adapt the discussion accordingly: it is not a “missing diagnostic term”, but some flux sent to the ocean during congelation. As it is congelation-related, it could explain the lower growth flux but should not directly affect the melt flux.

Thank you for drawing our attention to this interesting study, which we found particularly valuable in its clear presentation of the CICE mushy-layer scheme. The issue the authors identified with congelation growth not accounting for all basal energy loss may be contributing to the differences seen with these models, and we will reference it accordingly. However, we think that the main cause is something more fundamental.

Eq. (16) of Plante et al. may be particularly relevant:

$$\frac{\partial h_c}{\partial t} = \frac{F_{bot} - F_{cb}}{L\rho_i(1 - \varphi_{init})}$$

which expresses the rate of vertical congelation growth as the basal energy imbalance (numerator) divided by a term proportional to the solid ice fraction $(1 - \varphi_{init})$, which is small by default at 0.15.

What we think this equation will mean practically, is that only quite a small basal energy imbalance will be necessary to produce a given amount of ice *volume* production (due to the small denominator / large liquid fraction). Because ice volume production results in a shallowing of the vertical temperature gradient, and decreased ice conduction, we think this means that basal conduction will be naturally much lower in the mushy-layer scheme than in the BL99 scheme. The ‘missing term’ referred to is almost certainly the freezing of the liquid water entrained into the mushy layer. This appears to us to correspond to change in enthalpy of existing ice, and it is not clear to us that this energy budget term is systematically reported by the mushy-layer models, as it does not correspond to a change in elevation diagnosed by the standard CICE5.1.2 volume budget terms.

We will attempt to summarise this concisely.

- I find it a bit confusing that the selected CMIP6 members are referred to as the "IMB sample", given that they are compared to IMB buoys.

Yes, this nomenclature is maybe not the best. We think the reviewer's later suggestion of 'CMIP6 subset' is sensible.

- Figures are often mis-referenced in the text.

Apologies; we will endeavour to correct all instances of this.

Specific points and edits suggestions (some repeating the points above):

- L24: "although some improvement in agreement with reference datasets with model resolution and model complexity is discernible": This wording is not clear, please be more specific.

Most of the quoted studies find that higher resolution models, and models with more advanced sea ice physics, on average compare better to observations. We will make this both clearer and more specific – for example, by explicitly quoting which studies find this.

- L26-32: This paragraph is confusing, mainly because it is so concise that it becomes too vague. It would be worth expanding on these processes so that they are well understood by the reader before getting into the analysis. For instance, thicker ice melts more than thin ice: because of the larger area of ice surviving longer through summer?

We agree that the wording is confusing here, because we are probably unintentionally conflating two issues: seasonal sea ice melt, and sea ice loss in a warming climate. Seasonally, thicker ice melts *less* than thin ice, due to the albedo feedback. However, annual mean sea ice volume loss for a given increase in atmospheric forcing is on average greater if the initial sea ice volume is greater, due to the thickness-growth feedback (which primarily acts in the freezing season). This is the result we are referencing in Holland et al. (2006) and Chen et al. (2023). We will expand on this, along the lines indicated above, and ensure that these two processes are clearly separated.

- L30: it is not clear if "both" refers to the growth and melt, or to the processes.

This means both (ice volume) and (ice growth and melt). We will find a wording that clarifies this.

- L34: This sentence is confusing: it is not clear which climate variables you are referring to, and how it relates to the complexity of sea ice volume processes.

It is in areas such as this that a schematic would probably be most valuable. We will explicitly label those climate variables that are most important for sea ice evolution that we think have not previously been extensively evaluated.

- L39: “Although evaluation of the internal processes of the sea ice is in principle even more difficult”: This is a bit subjective, unless there is a reasoning added to this statement.

This is due to the even greater sparsity of observations of ice thermodynamics as compared to surface variables. We will clarify this.

- L42-45: This is also tedious to follow for readers not familiar with this study. I think it is worth adding some precisions.

We will expand on this with examples. One of the key findings of West et al. (2019) was that the ice thickness seasonal cycle of HadGEM2-ES was too amplified; the IMB evaluation showed both ice growth and melt to be much stronger in the model than in the buoy measurements.

- L65: why not simply “CMIP6 subset” instead of IMB? It would be less confusing when discussing models vs IMB observations.

Yes, this is better and we will make this change.

- L68: fast ice growth -> rapid ice growth (to avoid confusion with fast ice)

A good suggestion. Thank you.

- L77: You could cite Bitz and Lipscomb (1999) here.

Also a good suggestion; this will be done.

- Section 2.2.: Are all these IMBs all from CRREL with 10cm vertical resolution, or is there a number of them with higher resolution (e.g., SAMS, SIMB3, etc.)? It would be useful to indicate how they relate/compare to other data (for instance MOSAIC IMBs, etc., e.g. Koo et al., 2020)

Yes, these are all IMBs from CRREL with 10cm vertical resolution, because this was the dataset used in West et al. (2020). As this is primarily a model evaluation paper, we feel it would be out of scope to include new in situ observations, or use these to evaluate the IMB dataset. This is particularly true because converting the IMB data to a form usable for model evaluation was a very substantial task. Use of new buoy data would instead be a valuable study in its own right.

- L116: “was found to demonstrate well” -> displayed? Was used to characterize? It currently sounds like the quality of the observations were assessed against other unnamed references.

It would indeed be accurate simply to say that the IMB dataset displayed seasonal and regional variability, but we are trying to make a stronger statement: that the nature of this variability was consistent with evidence from other data sources. For example, the IMB data suggested a later onset of melting, and earlier cessation of melting, in the North Pole than the Beaufort Sea region; this is consistent with satellite measurements (e.g. Markus et al., 2009). As a second example, there were many more instances of nonzero winter ocean heat fluxes in the North Pole than the Beaufort Sea region; this is consistent with our understanding of the circulation of Atlantic Water within the Arctic Ocean, as the North Pole region is much closer to the AW inflow.

Both these examples are discussed in West et al. (2020), but would probably represent too much detail for the current study. We will expand a little on our point – perhaps say ‘The dataset displayed seasonal and spatial variability consistent with observational and theoretical understanding of the Arctic Ocean climate.’

- L121: Please add reference.

Apologies for this oversight, the reference is West et al. (2019) and we will add this.

- L125: Do you mean that you use both the sounders elevation measurements and the temperature profiles to determine the material interface positions? Also, I would like to have a measure of the uncertainties, and how it compares with other methods (e.g., see Richter et al., 2023)

The interface positions were determined using only the elevation measurements. However, there were multiple instances of decreases in surface or snow-ice interface elevation at times when surface melting could not conceivably have been taking place. Hence temperature measurements were used to validate the top melt fluxes: if a decrease in elevation occurred on a day when surface temperature was below a fixed threshold (-2C) the decrease was judged due to some process other than melting (e.g. wind drifting) and the melt flux for that day reset to 0.

In West et al. (2020) we assess uncertainties in IMB-estimated fluxes due to a number of factors, including salinity, conductivity, density and choice of reference layer. We did not explicitly discuss uncertainty due to direct elevation or temperature measurement uncertainty. However, we note that Lei et al. (2014) quoted an accuracy of 0.01m and 0.1K for elevation and temperature measurements from an IMB similar to those used in this study. These values would imply uncertainty more than an order of magnitude smaller than that due to those issues we do explicitly evaluate. We will briefly mention this.

This aside, Richter et al. (2023) was illuminating to read; it's encouraging to see that reasonably accurate ice thickness measurements can be made in the absence of elevation sensors.

- Section 2.3: Some of these data are also based on models, which are somewhat related to the components in some of the CMIP6 members. Could this interfere with the results? For instance, PIOMAS is, to my understanding, based on the POP model, which is also used in some of the CMIP6 members.

This is a good point, although we think that PIOMAS and ERA5 are the only datasets affected by this, and only PIOMAS includes an explicit sea ice model (ERA5 assumes all sea ice to be 1.5m thick).

- L143: "the sea ice state simulation of the IMB subset": rephrase. Perhaps: the sea ice state simulated by the subset models?

Yes, that sounds better. We will rephrase this.

- L144: "we restrict the evaluation"

Yes, this reads better and will be changed.

- L155-156: This could be presented with respect to the PIOMAS and CryoSat-2 seasonal cycle.

Thank you, we will mention as well when these reference datasets display minima and maxima.

- L170 suggestion: "There is strong correlation (0.81) between [...]"

Yes, this is better and we will make this amendment.

- L172-178: This paragraph is a bit difficult to follow. Some revisions would be helpful.

E.g.:

o “We compare the annual mean ice thickness to the anomaly in global 2m air temperature relative to the 1850-1899 average”.

Thank you for the suggested rewording, we will make this change.

o Clarify “Arctic Ocean temperature”: sounds like ocean temperature. I believe that you rather refer the T2m.

Yes, we do and will clarify this.

- L197: “Figure 2” -> I believe you are referring to Figure 1. Many other figures are also mis-referenced (e.g. Figure 7,8,9)

Yes, this is Figure 1. We will amend the misreferencing.

- Figure 4: Remove the “down” from the labelling of the SW down as it shows upwellingSW. It is also difficult to identify which curves is the net or downwelling radiation as they are not completely staggered (due to the CMCC curves). I recommend moving the net SW in a separate plot.

We will relabel these graphs ‘SW radiation’ and ‘LW radiation’, and make it clear in the caption that downwards=positive. We will try to make the net SW / downwelling SW plots more distinct.

- Figure 5: Missing information in the caption. The box plots are distributing the monthly means… for each year in the study period?

Yes, this is averaged across all years in the study period (1985-2014). This will be clarified.

- L246: typo :: or -> for

Thanks for noticing; this will be corrected.

- L263-267 (and also L501-502): I am not sure I understand the 2nd. Are you referring to the fact that IMBs will not sample the new ice forming in leads that form during the observation period?

It’s not so much the new ice **in leads** which isn’t sampled enough – though it isn’t – that would be more comparable with the modelled frazil flux. The problem is thin ice, already formed; this normally grows rapidly, is included in the modelled congelation flux congelation, but is by its nature also insufficiently sampled by the IMBs.

This is primarily an Eulerian-Lagrangian problem. Models report their diagnostics from an Eulerian perspective; they report on the state of all the ice within a specific area, and therefore automatically include the contribution of all rapidly growing thin ice. IMBs report from a Lagrangian perspective as they are advected around with the ice. Hence a rapidly growing new ice floe is sampled only very briefly, by its very nature; soon it thickens to a point at which it is no longer rapidly growing. It contributes only a small amount of rapid growth, or strong conduction, to a monthly mean basal growth or conduction flux.

Another way of thinking about this: even if the IMBs were **deployed** at random, they do not measure a random sample of the sea ice over time, because the probabilities of a particular ice type being sampled by a particular buoy at two different times aren’t independent. If at time t the sampled ice is thin and growing rapidly, the probability that the sampled ice is thicker and no longer growing rapidly is greater than the unconditioned probability for times greater than $t+\delta t$, for some relatively small δt.

- L290: "but display negative conduction in summer". This is interesting. Is this computed from the ice interior, or the conductive flux diagnostic? Is this indicating that the surface temperature is colder than the ice interior despite warmer air temperature?

All conductive fluxes come directly from the models' own reported diagnostics. In this case, a negative top conduction flux would indeed usually indicate that the ice interior is warmer than the ice surface. In the mushy-layer configuration of CICE5.1.2, the top conduction flux forms part of a fully implicit solution for the whole ice and snow column. It is proportional to the difference between the surface skin temperature and the top layer temperature, though the scaling conductivity depends on the (fully prognostic) salinity. We cautiously interpret that, on average, the top layer of the ice-snow column remains warmer than the surface throughout the summer in these models, and will note this in our revision.

It is a counterintuitive result, as atmospheric forcing might be expected to warm the surface to the melting point while the ice interior was still somewhat colder. However, the IMBs are agnostic on this point; some display weakly positive top conductive flux in summer, some weakly negative. It may be related both to the enabling of penetrative solar radiation, and the higher liquid fractions, in these models (such that there is less ice for the radiation to warm to the melting point).

- L310-311 (also L328-330): Please clarify : I do not get how a missing diagnostic term in the energy balance would impact the conduction computed from the simulated temperature profiles. However, as I mentioned above, I believe that you may be referringto the mistreatment of the energy balance in the mushy layer congelation scheme (not only a diagnostic), which results in a wetter and warmer ice base.

It is our understanding that the mushy-layer thermodynamic scheme is often characterised by weaker thermal gradients at the ice base (as energy exchange is inhibited by latent heat transformations), hence smaller basal conductive fluxes are to a degree not surprising – though we were taken aback by the magnitude of the difference.

When we refer to missing terms in the basal energy balance, what we are really referring to is the 'simple' basal energy balance equation (latent heat of fusion * d/dt ice base elevation = ocean heat flux + downwards basal conductive flux). This equation is valid for most thermodynamic ice models, but not the mushy-layer model – because d/dt ice base elevation does not completely characterise the latent heat exchanges that take place at the ice base. 'Missing terms' was possibly not a very good way to describe this. We will try to improve this.

(See also our response to general comment 4, which hopefully makes a similar point with different words).

- L344-345: In the CICE mushy layer scheme, a significant portion of the ice growth happens via frazil formation (DuVivier et al., 2022), again due to the treatment of the conductive flux in the congelation scheme. The fact that the frazil flux is not included here may thus impact the mushy group more than the other models.

We agree that this issue is likely contributing to the difference with these models and will state this, but think it is probably not the whole cause (see our replies to general comment 4, and to L388 comment below).

- Figure 7: In the panel c, the IMB points lie outside of the IMB uncertainty shading… Is that a plotting error?

Yes, the averages for the Beaufort Sea region have been mistakenly plotted here instead of the North Pole (while the shaded areas correctly represent the North Pole). We will correct this.

- L350: Is there more behind this attribution? It is not obvious when looking at Fig. 7 (which is also is referred to as Fig. 8 in the text). One could argue that the relationship is weak even among the model groups (e.g., there is low top melting in the purple models, without much of a slope).

Partly that is an effect of the vertical scale, such that the intermodel variability in top melting is quite hard to see. For example, the mushy-layer models have correlation between top melting an ice thickness of -0.99 (!!) in the North Pole region, which isn't obvious from the figure because the top melting variation is small on this scale. Among the GSI8.1 models correlation is -0.69, still substantially higher than the correlation across the ensemble as a whole (-0.51). The disconnect between whole ensemble and model group correlation is much reduced when looking at total melt rather than top melt, so we think that our point was valid – but probably it should be backed up with numbers given the difficulty of reading this from the graph.

- L358: "overlapping between ice growth and melt terms" -> rephrase: the terms are not overlapping, their season is.

Proposed rewriting: 'any ice growth or melt which occur in the same month are effectively invisible to the ice thickness seasonal cycle'.

- L375 (comment): I think that the fact that the CNRM-CERFACS models also display large conduction indicates that we have here a thermodynamic issue, rather than a diagnostic one (i.e., a problem in diagnostics would not influence the simulated internal temperature profiles).

This seems reasonable, but in the absence of further guidance from the model authors we will not speculate further. It is possible that a diagnostic issue might directly affect multiple output fluxes without affecting the internal thermodynamic evolution. Regardless, the conclusion that the diagnostics themselves are inaccurate is certain, and confirmed by the model authors.

- L388: It is also likely that lower growth is partly associated to the missing frazil contribution, which is more impactful using the current CICE mushy layer scheme.

With respect to the lower basal growth evaluated in Figure 5e,f this is almost certainly the case and we will state this. However note that in Figure 2c ice growth/melt is diagnosed as the amplitude of the ice thickness seasonal cycle; relative to ice thickness, this quantity is also quite low in the mushy-layer models. As calculated, it should include the frazil ice growth term as well as the congelation growth. So frazil/congelation splitting probably isn't the whole story with these models, though it is part of it.

- L389-391: I am not sure I got this right, perhaps it is worth spending more words here to clarify. I.e, the high (outward?) net LW flux is indicative of a cold surface temperature, despite the warmer atmosphere?

The net LW flux, and all radiative and conductive fluxes, are reported using the sign convention downwards=positive. Hence a higher net LW flux in one model than another indicates that the downwelling LW difference is relatively greater in magnitude than that of the upwelling component.

The downwelling component, very roughly speaking, diagnoses the state of the atmosphere, while the upwelling component, as it is determined by the surface temperature, is affected also by the state of the sea ice.

All other things being equal, we would expect an increased downwelling LW flux to alter the surface temperature and increase the upwelling LW accordingly. The relatively greater net LW in the mushy-layer models suggests that there may be a structural factor inhibiting this response to some extent. Reduced conduction through sea ice is one such plausible mechanism.

- L440: “pushing points to the left” -> towards smaller insulance ?

Yes, this is the meaning and we will state this.

- Figure 9: It looks like there is also vertical differences in the distributions. Are the conductive fluxes also computed differently in panel a and b?

No; for each grid point, the computed conductive fluxes are exactly the same. It is the thermal insulance that changes. Hence each insulance class is made up of a different set of points in panel a and b.

For example, a point that has an uncorrected thermal insulance in the 1.5-2 Km2W-1 range might then have a corrected insulance in the 1-1.5 Km2W-1 – so it would transfer from one distribution to another. This means that the bars are showing different distributions of conductive fluxes between panels a and b – because they are made up of different sets of points, even though each individual point has the same conductive flux.

We will try to improve the wording here to make this clearer.

- L466-475: I feel like there is a missing point here: to me, this analysis is indicating that the relationship between ice thickness and top-melt is a large scale one (i.e., it related to the ice Area, via the albedo effect), and thus is not showing when looking at individual grid points and IMBs.

Yes, this is a good point and we will state this. We believe it does not invalidate the main point of this paragraph, which is that the sampling bias of IMBs does not significantly impact the measured top melting fluxes – indeed, the melt-thickness correlation being less visible on the small sale of the IMBs is one reason for this.

- L476: ameliorate -> address?

Better. We will make this change.

References:

Bitz, C. M., and W. H. Lipscomb (1999), An energy-conserving thermodynamic model of sea ice, J. Geophys. Res., 104(C7), 15669–15677, doi:10.1029/1999JC900100.

Koo, Y., Lei, R., Cheng, Y., Cheng, B., Xie, H., Hoppmann, M., Kurtz, N.T., Ackley, S.F., Mestas-Nunez, ˜ A.M., 2021. Estimation of thermodynamic and dynamic contributions to sea ice growth in the Central Arctic using ICESat-2 and MOSAiC SIMBA buoy data. Remote Sens. Environ. 267, 112730. https://doi.org/10.1016/j.rse.2021.112730.

Plante, M., Lemieux, J.-F., Tremblay, L.B., Tivy, A., Angnatok, J., Roy, F., Smith, G., Dupont, F., (2024), Using Icepack to reproduce Ice Mass Balance buoy observations in landfast ice:

improvements from the mushy layer thermodynamics, The Cryosphere, 18, 1685–1708,https://doi.org/10.5194/tc-18-1685-2024.

Richter, M. E., Leonard, G. H., Smith, I. J., Langhorne, P. J., Mahoney, A. R., and Parry, M.: Accuracy and precision when deriving sea-ice thickness from thermistor strings: a comparison of methods, J. Glaciol., 69, 879–898, https://doi.org/10.1017/jog.2022.108, 2023

References for review response

Lei, R., Li, N., Heil, P., Cheng, B., Zhang, Z., and Sun, B.: Multiyear sea ice thermal regimes and oceanic heat flux derived from an ice mass balance buoy in the Arctic Ocean, J. Geophys. Res.-Oceans, 119, 537–547, https://doi.org/10.1002/2012JC008731, 2014.

Markus, T., J. C. Stroeve, and J. Miller (2009), Recent changes in Arctic sea ice melt onset, freezeup, and melt season length, J. Geophys. Res., 114, C12024, doi:10.1029/2009JC005436.

West, A., Collins, M., Blockley, E., Ridley, J., and Bodas-Salcedo, A.: Induced surface fluxes: a new framework for attributing Arctic sea ice volume balance biases to specific model errors, The Cryosphere, 13, 2001–2022, https://doi.org/10.5194/tc-13-2001-2019, 2019.