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.

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.

- 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).

- 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.

- 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.

- 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.

- Figures are often mis-referenced in the text.

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.

- 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?

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

- 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.

- 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.

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

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

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

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

- 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)

- 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.

- L121: Please add reference.

- 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)

- 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.

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

- L144: “we restrict *the* evaluation”

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

- L170 suggestion: “There is strong correlation (0.81) between […]”

- L172-178: This paragraph is a bit difficult to follow. Some revisions would be helpful. E.g.:
  - “We compare *the* annual mean ice thickness to *the anomaly in global 2m air temperature relative to the 1850-1899 average*”.

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

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

- Figure 4: Remove the “down” from the labelling of the SW down as it shows upwelling SW. 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.

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

- L246: typo :: or -> for

- 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?

- 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?

- 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 referring to 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.

- 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.

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

- 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).

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

- 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).

- 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.

- 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?

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

- 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?

- 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.

- L476: ameliorate -> address?

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.