**Containing Author Comments**

**Reviewer 2**

Peer review of "Data-driven emulation of melt ponds on Arctic sea ice" by Driscoll et al.

The manuscript presents a Machine-Learning (ML) model intended for estimating melt-pond fraction on sea ice based on meteorological variables and developed and tested against spectrometer satellite data of Melt-Pond Fraction (MPF). The work appears solid but the manuscript is somewhat short on methodology and review of earlier work. In addition, the manuscript seems to claim that melt ponds have not been taken into account in climate models before, which is not the case. Here follows some mostly minor comments and suggestions that the authors can take into account before acceptance for publication is considered.

We acknowledge the reviewer's comments and have been very grateful for them. We believe going through them and adding these details and insights help yield a substantially improved manuscript.

Comments and suggestions

1. In particular in the abstract line 5, the manuscript seems to claim that melt ponds are not included in GCMs, and that melt ponds therefore play no role in IPCC future projections of sea ice. This appears not to be the case, indeed many models have included parameterisations of melt ponds, see e.g Sterlin et al, Ocean Modelling, 2021, and references therein, as well as Roeckner et al., Journal of Advances in Modeling Earth System, 2012, and Diamond et al, Journal of Climate, 2023.

We thank the reviewer for this comment. We apologise for the confusion we have caused. We intended not to imply that melt ponds are not used in GCMs or projections in any way, but that historically it has been often the case that GCMs are run without them, and that thus key processes for sea ice prediction are likely being missed. (A similar conclusion is given in Diamond et al. 2023 itself who point out that only 11% of CMIP6 models have an explicit melt pond scheme).

Our message was intended in the spirit of Diamond et al. 2023, but not clearly communicated by us, which we have updated. We also included these and multiple other references demonstrating where indeed the inclusion of melt pond processes shows them to have an important role on sea ice prediction. We have hopefully given a useful account therefore of the literature provided here, as well as other references.

In addition to correct this aspects of historical uses of melt-pond parameterisations in ECMs, the manuscript should also provide a much more extensive review of historical efforts to provide melt- pond parameterisations, including for instance reviews of Lühtje et al., J. Geophys. Res – Oceans, 2006, and Flocco and Feltham, J. Geophys Res. - Oceans, 2007.

We have updated our manuscript to include a detailed history review of melt pond parametrisations and approaches. We thank the reviewer as it substantially improves our paper.

2. Test against SHEBA. In order to test performance of the developed ML model with other types of data and to results from other parameterisations (e.g. Holland et al, J. Clim., 2012), the SHEBA data may be applied.

We thank the reviewer for this comment. The SHEBA campaign has represented a valuable dataset in the sea ice literature. Our study was designed around a different objective: the spirit of our model is to inherently create an emulator that is created and tested against broad-scale, long-term, pan-Arctic observations. The MERIS and OLCI datasets therefore are ideally suited for training and evaluating a data-driven emulator intended to generalise across the Arctic domain.

In this work we chose to evaluate our emulator's performance as whole against pan-Arctic data, which we feel is a strength. Other data have merit, but they have spatial and temporal limitations (thus can be affected by processes specific to one location and conditions affecting one year). These do not necessarily provide a representative testbed for the generalisation we wish to achieve. By demonstrating one can emulate directly from observations we feel this is a key first step, and we have included many potential future avenues in the paper now. We have highlighted the reviewer's point and included our motivational reasons much more in our text.

3. Section 2.1 first paragraph: Expand the description of the MPF data, e.g. regarding how MPF is obtained from these data. Are there missing data and due to what?

We have added in details on the MPF data, how it is obtained, missing data and its sources accordingly.

4. L116-119: The conclusion regarding change in pattern between the two observational data set beibg due to instrumental differences or climate changes would be better explored if training data were chosen for only MERIS, 2002-2011.

We have explored this by using training data that covers the MERIS only period of 2002-2011. Indeed the R2 score drops from 0.63 to 0.54 and the MSE increases from 0.0030 to 0.0038 (an increase of approx 27%). So we see that a MERIS only trained emulator causes a greater separation between observations and emulator in the second satellite era. That there still exists a good predictability when trained on MERIS data only to predict the other is still a positive result, but this test reduces our ability to attribute shifts to climate change. This has shaped our narrative and also motivated the inclusion of results from an additional melt pond fraction dataset where we see very good emulation scores. We have updated the conclusion to reflect the nuance above. We thank the reviewer for the suggestion and this test that has enlightened the discussion.

5. First two paragraphs of section 3.2: Please more extensively describe the procedure of constructing the model, including finding the hyperparameters, in a way that none ML experts can follow. Shortly describe all “ML words”. “Plateaux” → “converge”.

We thank the reviewer for this. This has been now done.

6. L69-72: Please first describe the earlier studies before indicating differences to yours.

Our work is substantially different and we have clarified this now – for example Peng et al. 2022 use neural networks and statistical temporal filters to fill observational gaps, and effectively interpolate over missing or obscured melt pond data so as to ensure continuity. In essence their work reconstructs missing melt pond data, using a statistical and ML interpolation system. Ours seeks to learn the physical mapping from climate data to melt pond fraction. Thus our work importantly avoids statistical filling, uses only direct observational data and instead seeks to build relationships to physical variables. We have added additional information to clarify this and therefore show how our study is distinct.

7. Fig. 1 caption: Explain how the standard deviation is calculated, what are the input data?

We have explained this in the caption.

8. L124: Write out the abbreviation “DMIOI-L4”.

Done.

9. Fig. 2 caption: In the first sentence before “note that” indicate the period averages are taken over. The “note that” part is quite awkwardly formulated, perhaps “Note that if at least one observation exists for a given grid point this point is included”.

Amended. Thank you.

10. L135-137: The sentence is awkwardly formulated.

Thank you for pointing this out. We have rewritten this for clarity and flow.

11. L155: What is “level ice”?

The level-ice means not ridged, or deformed ice, and the associated ‘level-ice’ melt pond scheme developed by Hunke et al. (2013) is designed to represent the core concept that meltwater collects predominantly on level-ice, not on ridged or deformed ice. We have added this explanation to our text.

12. L189: “unseen” seems to be exaggerated given that The ML model build on observational input. Perhaps it suffice to conclude that the model can reproduce MPF associated with stochastic weather variability.

Apologies for our confusing wording– we intended “unseen” to refer to the fact it is test data. We agree and have reworded it so it now refers to capturing physical events in the Arctic associated with stochastic weather variability (in the test data).

13. Fig. 5: Why not show root-mean square error (RMSE) of model results? Perhaps also show anomaly relative to climatology of observations, in order to compare (and hereby regard climatology as a persistence model).

We have amended the figure to have alongside it the values of RMSE for easy comparison and thank the reviewer.

14. L197: What is “R2 score”?

We have added in an explanation of R2 score.

15. Fig. 6: In (b) the scale of the shading is not indicated.

We have added information now.

16. L255-259: This paragraph is not so clear.

We believe we have clarified this now.

Typos:

L92: “data is” → “data are”, data are plural. Same in Fig. 2 caption. And Line 235. L96: Remove “one”.
L130: “of”->”as”.
L245: “it is” → “that it is”.

We have amended all typos, except for keeping L235 as a mass noun.