Containing Author Comments

Reviewer 3

Review comment for 'Data-driven emulation of melt ponds on Arctic sea ice' in *The Cryosphere*

In this study, Driscoll et al. present a point-wise emulator for Arctic melt ponds, developed using pan-Arctic satellite observations and reanalysis data. They demonstrate its ability to predict Melt-Pond Fraction (MPF), which shows promise for integration into global climate models (GCMs).

Given the increasing prevalence of melt ponds due to ongoing global warming, research on melt ponds in polar regions is becoming increasingly critical in the context of climate change. However, while the study provides valuable insights, there remains considerable room for improvement.

Therefore, I recommend that the manuscript undergoes major revisions, and specific comments are outlined below.

We are extremely grateful for the reviewer's comments and insights, and by incorporating them it has improved our manuscript substantially.

1. The introduction section would benefit from further elaboration, particularly with regard to the context and rationale behind the study. The following points, though not exhaustive, serve as examples:

(1) Lines 65-69: Driscoll et al. (2024) have already developed an emulator for melt ponds based on model data. If this existing emulator performs adequately, what is the justification for advancing to an emulator built solely on observational data? A clearer explanation of the advantages and motivation for this shift would strengthen the rationale for the current work.

Thank you for this point. Our previous emulator was trained on model-generated melt pond data and therefore inherited the structural assumptions and sensitivities of the underlying scheme. However, by demonstrating that such an emulator could run stably within a model, we showed the feasibility and potential value of using emulators - particularly if one could link climate variables directly to observational data.

Yet transitioning from model data to observational data in emulators has often been a prohibitive step in climate research. In contrast, the current study does just that and therefore demonstrates the next step: it learns from observed melt pond fraction, aiming to reduce uncertainty rather than reproduce it. We have clarified this in the revised text. We have also highlighted the significance of a first step in making an observational emulator in the context of wider climate research.

(2) Lines 70-72: The distinction between the approach used by Peng et al (2022), which involves "interpolating melt pond observations across a temporal series," and the methodology employed in the current study, which involves "regridding satellite data of melt ponds," needs further clarification. Specifically, how do these two techniques differ in terms of their methodology, accuracy, and overall Applicability?

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 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.

2. The methodology section would benefit from additional explanations to better justify the importance of the chosen approach. The following points, though not exhaustive, are provided as examples:

(1) Lines 102-103: Given the significant differences in the amplitude of the seasonal cycle and temporal variability, how might these factors affect the consistency of the data between the training and test periods? Additionally, considering that Arctic sea ice is expected to increasingly resemble MIZ conditions (Line 242), would using segmented training data lead to improved

results and potentially alter conclusions, particularly with regard to the identification of feature importance?

The slight difference makes our goal harder, thus we believe our emulator is doing well. We have also accommodated the reviewer's insight here by testing on another version of this dataset where the OLCI peaks do not suffer from this discrepancy: our performance (with a second trained emulator) improves. Success across two datasets adds to the robustness of our approach and the validity of our pipeline. We have added the details of this work and our scientific findings into our paper. We thank the reviewer for spotting this and therefore enriching our paper through this work.

We acknowledge the reviewer's second insightful point. Whilst the goal of our paper is to identify useful observational variables and build a pointwise emulator from observed inputs, such that it is broad-scale, long-term and pan-Arctic, it is absolutely possible that one might be able to build a "MIZ only" emulator in the future. It is likely that with distinct physical processes in the MIZ, new features would be required to emulate this well – for example those that represent interconnectivity in ponds, surface roughness and floe topography given a younger, flatter, yet more broken surface. There are multiple substantial challenges. Given greater uncertainties around quality of different MIZ data one would test many different data sources, campaigns and imagery to capture desired effects required by potentially new and necessary variables from such datasets (perhaps developing tools to super-resolve existing data). We have in our conclusions noted other challenges that come with this. For example, there is no single definition of the MIZ and such a study could include and test many definitions. We are in consultation with groups that have access to proprietary datasets for such purposes and are considering a research proposal to similar effects.

We are excited by the reviewers comment and have incorporated the reviewer's point both discussing the impacts of feature analysis that a "MIZ only" focus might have (including those with new necessary variables), as well as discussing now in more detail MIZ emulators as a potential future research direction. Ideally our study serves as a key first step – one that has often eluded other areas of climate research in building observational emulators – and helps set up future investigations including that one.

(2) Line 138: The description of the specific training process is overly simplistic and fails to capture the nuances of the neural network architecture employed in this study. A more detailed explanation of the training procedure, including the rationale behind the choice of neural network model, would enhance the clarity and rigor of the methodology.

We have added more detail behind the choice of neural network model, and the training involved to enhance clarity and rigor. We thank the reviewer for this point.

3. The results section would benefit from additional discussion to emphasize the significance of this study and its contributions to the field (1) Line 156: Can the conclusions regarding feature importance drawn from the model data be directly applied to a study based solely on observational data? It would be valuable to discuss the potential limitations or caveats when transferring findings from model-based to observational datasets.

We thank the reviewer for this. Indeed we believe it to be a significant step and we have added additional discussion on such significance – such as the difficulties faced in other areas of creating observational emulators as compared to synthetic data without gaps and noise.

We appreciate the second point fully. And we have added in a greater discussion on such caveats.

(2) Lines 164-165: Considering that Arctic sea ice is becoming thinner and more fragmented (Line 242), advection is expected to play an increasingly significant role in the evolution of melt ponds. However, this study only considers the influence of wind speed, while neglecting the impact of wind direction. Could the omission of wind direction potentially affect the study's conclusions, particularly in light of the changing dynamics of Arctic sea ice?

We thank the reviewer for this point. We therefore tested the emulator with and without wind direction (i.e. not only the magnitude) and the results are essentially identical including in the MIZ. However there was an extremely small, almost negligible, improvement by using the magnitude. Therefore we kept magnitude, and our conclusions are not affected. We have now added these details to inform and enrich our paper. We thank the reviewer for raising this point.

4. Several typographical errors and omissions need to be addressed: (1) Figure 1: The distinction between observational data and model results is not clearly indicated in the image. It would be helpful to specify how these two data sources can be differentiated visually.

We apologise for this – this is indeed a typo and has been updated and clarified.

(2) Lines 172-173: Further clarification is needed regarding the "bandit based approach." A more detailed explanation of this method would enhance the reader's understanding of its relevance and application in the current study.

Done.

(3) Line 182: The "optimiser Following" should be corrected to "optimizer. Following."

Corrected.

(4) Figure 5: The figure legend and the image itself do not align. Additionally, the year should be added to the x-axis to provide clear temporal context for the data.

We have updated the figure accordingly.

(5) Line 196: The definition of R2 should be provided for clarity, as some readers may not be familiar with the metric or its specific interpretation in this context.

We have provided a definition of R2 score.