Reviewer Comments

1 Summary

In this study the authors introduce an ML based model, PIXAL. It implements an XGBoost-based architecture that maps 14 environmental features from the MAR climate model to Greenland ice albedo from MODIS observations. The architecture significantly outperforms traditional parameterizations ($R^2 = 0.563$ vs. 0.062). The strength of the study lies in the application of SHAP, an interpretability tool, to PIXAL. The SHAP analysis yields surface height and temperature as the primary drivers of the change of ice albedo changes. While providing a robust data-driven alternative to MAR formulations, the model operates as a feature-based regressor. Although suitable for hindcasts and forecasts, PIXAL can struggle with out-of-distribution samples in a changing climate. I think this is a valuable study and hope that my comments and technical critiques are useful to the authors in further strengthening the manuscript.

2 Reviewer Comments

2.1 Major

- **The physics-informed aspect of PIXAL**

  As a general comment, I would encourage the authors to clarify exactly the "physics-informed" aspect of the ML model developed in the study. Physics-informed models typically use knowledge from physics to introduce constraints or loss functions to nudge neural networks to produce outputs that are consistent with the physics of the problem at hand. Could the authors perhaps clarify their claim and/or sources as to what exactly makes their model physics-informed?

- **Use of a tree-based model architecture**

  - **Generalizing capabilities**: PIXAL is based on XGBoost, which is a tree-based ML model. It is widely understood that tree-based methods can have issues with generalizing outside of their training dataset. Do the authors think using a neural network based model architecture be more useful for generalization? Is there a particular reason/advantage for choosing a tree-based model in this case?
  - **Spatial structure**: The ice albedo is a spatial feature. XGBoost does not understand the relationship between a pixel and its neighbors and from my understanding, in this study each grid is treated as an independent data point. Can the use of a convolution based method help capture the gradients in the dark zone more accurately?

– **Scalability**: The authors recognize that the unexplained variance in the ice albedo prediction can also be attributed to a lack of information about several physical processes. For XGBoost, adding more features (such as information about algae and dust, even if contrary to reality, they were available atleast as proxies) would significantly increase the computational cost. Have the authors considered other methods of representing this information without coming up against the difficulty of computational scalability of tree-based models?

2.2 Minor

- Line 37: ”hindrance to” instead of ”hindrance for”
- Line 88: ”underestimation” instead of ”underestimates” fits better
- Line 213: more appropriate in ML terminology would be ”generalize to data”
- Line 229: ”predictand” instead of ”predictant”
- Line 236-240: Could these be put into a table? If the authors wish to leave the hyperparameter information in a paragraph, some of the text here seems to be in a different font.
- Line 272: ”insufficient variability” instead of ”too-low variability”