Review 2 of Antwerpen et al 2026, TC Discussion

Note to editor and referees: we decided to revise our previous replies, please use the following responses.

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 (R2 = 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.

We would like to thank Referee #2 for their thorough comments and helpful suggestions. We appreciate the effort it took to review our manuscript and are confident the manuscript will be improved after we address their comments.

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?

Thank you for your comment. I agree with your statement as we did not include any physical laws into the model or loss function besides setting a constraint for the predicted ice albedo to be in the range 0-1. I changed the title, abstract, and introduction to reflect this with “Physics-Inspired” instead of “Physics-Informed”.

Use of a tree-based model architecture

Generalizing capabilities: PIXAL is based on XGBoost, which is a treebased 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?

Thank you for raising this point. Our initial tests for ML model selection included neural networks (NN) as well as tree-based approaches. We found a lower performance for NNs and decided to continue with the best-performing models and omit the NNs from further tests and analyses. In L462-469 we describe that our dataset has a ~6-sigma spread in values, ensuring some robustness against future out-of-distribution values. Please also see our response to the next comment, where we discuss the model architecture choices (including NNs) and add additional text to the limitations in the discussion section of the manuscript.

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?

Thank you for this suggestion. Similar to the comment above, our initial tests for ML model selection included a convolutional neural network (CNN) and a CNN+long short-term memory (CNN-LSTM) model. These spatially-aware architectures showed better performance than a standard NN, though still not as good as XGBoost. We acknowledge that SHAP can be applied to tree-based models, CNNs and NNs. However, implementations of SHAP for CNNs and NNs rely on approximations of SHAP values, while SHAP for tree-based models computes the exact SHAP values. Given the importance of the explainability of the model in this manuscript, and the superior performance of XGBoost over NNs and CNNs, we opted for XGBoost as the basis for PIXAL.

We agree that the generalization limitations (compared to NNs) and spatial limitations (compared to CNNs) of our tree-based approach may affect the results and that this requires acknowledgement in the manuscript. We therefore added the following paragraph to the end of the limitations section in the discussion:

"Beyond these distributional considerations, we make two architectural trade-offs by opting for XGBoost as the basis of PIXAL. First, neural networks can in principle generalize outside of the training distribution, whereas tree-based models will assign out-of-distribution inputs the same prediction that is associated with the nearest boundary of the training data, effectively limiting extrapolation. However, as described above, the ~6σ spread of our training dataset partially mitigates this limitation. Moreover, preliminary experiments showed a lower performance for NNs than tree-based approaches in modeling ice albedo. Second, tree-based models treat each grid cell as an independent data point and do not encode spatial relationships between neighboring pixels. This may limit their ability to capture spatial gradients such as those in the dark ice zone. Convolutional neural networks (CNNs) are theoretically better suited to leverage the information surrounding each pixel due to the spatially-aware nature of their convolutional kernels. However, our preliminary tests showed a superior performance of tree-based models over a CNN and a CNN-long short-term memory (CNN-LSTM) model in modeling ice albedo. Lastly, while SHAP can be applied to NNs and CNNs, these implementations rely on approximations of the SHAP values, whereas the SHAP implementation for tree-based models computes the exact SHAP values. Given the importance of model explainability to identify drivers of ice albedo variability, we opted for XGBoost as the basis for PIXAL."

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 at least 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?
Regarding computational scalability, training runtime does increase with additional input variables. However, in principle, training only needs to be performed once. Future iterations of PIXAL will likely prune non-informative variables, further reducing computational costs.

Regarding your point on adding additional LAC-related datasets or LAC proxies, we considered adding additional datasets to the input data collection. However, we found that no datasets are available on the same spatiotemporal range and resolution as the rest of our data (MAR and MODIS). We included the 2-year annual algal bloom dataset from Wang et al, (2020) in initial tests and found that the uneven input data availability resulted in biases and artefacts in our model output. We are unaware of other, more comprehensive datasets or proxies that are able to add sufficient information to improve the performance of the XGBoost model. To address this point, I added the following text to Section 5.2:

"Some datasets are available that could represent these LAC-driven darkening processes, e.g. annual algal bloom data (Wang et al., 2020). However, this data is currently not available with the same spatiotemporal range and resolution as the MAR and MODIS datasets used in this study. Initial experiments showed that this 2-year annual algal bloom dataset created uneven input data availability resulting in biases and unwanted artefacts in the model output. We therefore omitted this data from our input data collection. We are unaware of other, more comprehensive datasets that are able to add sufficient information to improve the performance of the XGBoost model. Beyond adding direct observational inputs, alternative approaches to representing LAC-driven darkening processes could include proxy variables that co-vary with LAC concentration, or explicitly accounting for the contribution of unknown processes through uncertainty quantification. However, identifying reliable proxies at the required spatiotemporal resolution is an open challenge. We consider both directions valuable for future development of PIXAL."

2.2 Minor
Line 37: "hindrance to" instead of "hindrance for"
Done.

Line 88: "underestimation" instead of "underestimates" fits better

Done.

Line 213: more appropriate in ML terminology would be ”generalize to data”
Done.

Line 229: ”predictand” instead of ”predictant”
Done.

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.
We prefer the text to be in a paragraph. I changed the font to match the rest of the manuscript.

Line 272: ”insufficient variability” instead of ”too-low variability”
Done.