

Author's Response

Summary of major Changes

Based on the three comments we made the following main changes to our manuscript:

- 1) We added a systematic input selection study on the atmospheric variables to determine which variables are necessary and which are redundant, complementing the optimization of the network architecture (sections 2.2.4 and 3.2)
- 2) We thoroughly revised the introduction and expanded the related literature, to highlight how our work differs from existing studies, making the novelty of our approach more visible.
- 3) We revised the data subsection in the methodology section to clarify the DMIHH model, which our ML model is based on (section 2.1.1).
- 4) We explained ML terminology and practices in more detail, including a clear explanation and justification of our train/validation/test split (section 2.1.3 and Appendix A).
- 5) Add additional result maps of some daily predictions throughout the year (Appendix E), as well as seasonally aggregated maps (Figures 6, 7).

Author replies to Anonymous Referee #1

General comments

This paper introduces a newly developed neural network-based emulator that predicts the temporal evolution of Greenland ice sheet surface melt. The emulator was trained on the output from the polar regional climate model HIRHAM5 and its firn model DMIHH, forced by the ERA-Interim reanalysis. It is clearly shown that the Modular NN configuration of the emulator, the standard setting developed in this study, can provide realistic information on the spatiotemporal evolution of ice-sheet surface melt, along with the daily melt amount. My impression is that this is a unique study that can provide useful information on the synergy between machine learning and cryosphere science. Although I think the information provided, in particular on the methods, can be improved, the results and discussion sound reasonable and sufficient to me. Therefore, I suggest that this paper can be published after revisions. I list some specific comments below.

A: We thank the reviewer for the positive and constructive assessment of our manuscript and appreciate the recognition of the novelty and relevance of combining machine learning with cryosphere modeling. We specify how we addressed the comments below.

Specific comments

L. 9 “mean absolute error below 0.23 mm w.e.”: Compared to what? What is the reference data for this comparison? Please explain.

A: We specified: “relative to the surface melt produced by DMIHH”.

L. 45 ~ 58: It is worth reviewing and citing the paper by Hu et al. (<https://doi.org/10.5194/tc-15-5639-2021>) in this part.

A: We thank for bringing this paper to our attention; we included it in our related work in the introduction.

L. 59 “high temporal variability”: Can the authors explain this point quantitatively and add a reference for this argument if possible?

A: We clarified that “high temporal variability” refers to the strong day-to-day fluctuations in melt at daily resolution compared to temporally aggregated (monthly/annual) quantities (line 73 in the revised version).

L. 60 “temporal context”: I don’t think this technical term is widely recognized in the cryosphere community. Can the authors introduce additional explanations about the term so that more readers can easily understand?

A: We now directly refer to “temporal information” or “history” to avoid technical terminology.

L. 60 “While the models predicting annual ~”: Do the authors mean that the models refer to “ML” emulator? Or RCMs? Please clarify.

A: We specified that we refer to the existing works (line 64 in the revised version).

L. 63: What do the authors mean by “lag effects”? Please explain in more detail.

A: We mean that surface temperature and melt onset depend not only on current atmospheric forcing but also on the previous thermal state of the surface and energy input

to the surface, introducing a short-term memory. We reformulated this section and state that surface melt also “evolves gradually” (line 67 in the revised version).

L. 67: What do the authors mean by “model generalization”? Please explain.

A: We specified that we mean generalization of the model to different locations and climate states (lines 61/61 in the revised version).

L. 73 “Our model can be re-trained on data for future scenarios ~”: If the NN will be used for the future simulations of the ice sheet surface melt, do the authors have to train the NN using the output from the future climate simulations by an RCM such as HIRHAM5? Please explain more explicitly.

A: Yes, applying the emulator to future climate scenarios requires retraining on RCM simulations produced under those specific forcings. Since the neural network learns a mapping from atmospheric input to firn model output, reliable extrapolation beyond the training distribution cannot be assumed. Validity of temporal extrapolation or interpolation (e.g. to a lower SSP scenario) need to be investigated and are out of scope for this work. We consider it essential to clarify this point in the introduction to avoid suggesting that the emulator is universally applicable to all HIRHAM atmospheric data, irrespective of the applied forcing. The issue is briefly revisited in the conclusions (lines 490-494 in the revised version).

L. 78 ~ 79: Please explain all the properties included in the daily output of the polar RCM HIRHAM5 with its firn model DMIHH.

A: We significantly changed the data description, highlighting the processes and relationships relevant for designing the ML model (subsection 2.1.1)

L. 79: What is the total snow and ice model layer thickness that DMIHH considers with the 32 model layers?

A: The 32 layers represent a firn/ice column of 60m w.e. thickness in total (mentioned in line 99 in the revised version).

L. 84: It is better to explain how bare ice is determined in the DMIHH model.

A: The firn model scheme DMIHH operates on all grid cells that are defined to belong to the ice sheet by an ice sheet mask. Bare ice is defined as the top layer having zero snow fraction (added in line 107 in the revised version).

L. 85: Atmospheric forcing for what? For DMIHH? Or for the newly developed emulator? Please clarify. In addition, please list all the properties included in the atmospheric forcing.

A: We now clarified that the whole subsection 2.1.1 explains only the physical model DMIHH.

L. 90: It is unclear what the “input data” are. Input data for DMIHH? Or input data for the emulator?

A: We now clarified that the whole subsection 2.1.1 explains only the physical model DMIHH.

L. 97 “they can be problematic when training ML models.”: Please explain the reason for this argument in more detail.

A: We added additional explanation in lines 122-123 in the revised version.

L. 107: Does the negative sensible heat flux mean that the heat flux directs from the ice sheet surface to the atmosphere? Or opposite? Please explain.

A: In HIRHAM5 the sign convention for heat fluxes is defined such that positive values indicate energy directed towards the surface, while negative values represent energy transfer from the surface to the atmosphere. Thus, the extreme sensible heat flux down to -400 W/m^2 corresponds to intense surface cooling, causing the surface temperature approaching 0 Kelvin (explained in line 134 in the revised version).

L. 129: Why is the number 5000 selected here? A more detailed explanation is needed.

A: We will add a justification for selecting 5000 grid cells, explaining the trade-off between computational feasibility and maintaining sufficient spatial representation across melt zones, as well as noting that this number provided stable training performance in preliminary experiments. We added additional explanation in Appendix B.

L. 178 “we choose the hidden layers of the network to be 64-128-128-64-32-16-16”: Please explain the meanings of each number, in particular for non-specialists in NN.

L. 182, L. 184, and L. 185: Same as the comment on L. 178.

A: We included a general introduction to neural networks as subsection 2.2.1 in the revised version, and introduce specifically the notation for nodes per layer in line 206.

L. 188: Please explain in more detail about “LeakyReLU activation function.”

A: We added additional explanation in lines 207-209 in the revised version.

L. 193 “the optimal number of days to be used in the short-term module”: What do the authors mean by “optimal”? Please explain in more detail.

A: In this setting “optimal” refers to the number of preceding days yielding the lowest validation error during hyperparameter tuning, and that performance gains became marginal beyond a certain number of days. We now reformulated: lines 247/248 in the revised version.

L. 215~216 “the total computational cost remains far lower than physical firm models”: Can the authors add quantitative information for this explanation? I think such information is useful for other emulator developers.

A: DMIHH needs about 2,5 hours per simulation year on 16 CPU cores. We added this information in lines 297/298 in the revised version.

Technical corrections

L. 89: It is better to add something like “within DMIHH” at the end of this sentence.

L. 111: It is better to add the mathematical symbol “ x ” after “heat flux values.”

Table 1 caption: Please add “Autoreg” after “the autoregressive element.”

Figure 3 caption: It is better to explain the numbers in Gt listed in each panel.

L. 319: Suggest adding “surface” before “atmospheric variables.”

A: We included these technical corrections in the revised version.

Author replies to Anonymous Referee #2

Overall, I think the authors have done a very good job. The manuscript is clearly written, the structure is logical, and the figures are generally of high quality. With some minor revisions, I believe the paper should be ready for publication. Please see the in-text comments for specific suggestions to improve clarity and to make certain figures more rigorous.

A: We thank the reviewer for the positive and constructive assessment of our manuscript, and the kind words. We fully understand the concerns regarding model robustness and impression of cherry-picking a specific test year and a specific day to illustrate model performance. Below we describe how we addressed these concerns in the revised version.

One point I would like to raise concerns the choice of using a single year (2016) as the test set. The manuscript does not provide a clear justification for this decision. Using only one test year raises the concern that the evaluation may depend on a particularly “lucky” year, or alternatively on a year with atypical behavior. In either case, it becomes difficult to convincingly demonstrate the model's generalisability.

As a reader without specific expertise in the Greenland Ice Sheet (GIS), I am unsure whether 2016 is representative of typical conditions or whether it may have experienced unusual or extreme events. From a scientific robustness perspective, it would strengthen the study to repeat the prediction-versus-observation scatter plots for one or two additional test years. This would allow the reader to assess whether, for example, the melt overestimation by the autoregressive and modular neural network models is a persistent feature or specific to 2016.

If the authors intentionally selected 2016, I would encourage them to provide a clear justification. For example, an appendix figure showing the distribution of SMB over the GIS compared to other years could help demonstrate whether 2016 is representative or exceptional.

A: We appreciate this important comment and want to explain our selection process for the test year. We note that evaluating one additional single year may not necessarily provide stronger evidence of generalizability, since performance on any individual year may itself be influenced by natural interannual variability.

We added further explanation on our choice of the test year in lines 141-161 and Appendix A.

A similar concern applies to Figure 5, which focuses on 21 July 2016. While this date is interesting, showing only a single summer day risks giving the impression of a carefully selected (“lucky”) example. I would encourage the authors to include additional dates in the appendix, ideally covering different seasons, for example, shoulder seasons or winter periods when little or no melt is expected. This would provide a more comprehensive picture of model behavior across varying surface mass balance regimes.

[A: We now added additional examples in Appendix E in the revised version.](#)

In addition, it could be helpful to show aggregated diagnostics, such as spatial plots of the mean SMB over several months (or seasonal averages) for the best-performing model. Such analyses would provide stronger evidence that the model captures robust patterns rather than performing well on isolated dates.

[A: We added Figures 6 and 7 in the revised version, showing seasonal aggregates of the test period and the whole data period.](#)

Overall, I consider this a valuable contribution, but addressing these points would further strengthen the manuscript.

Author replies to Editor Comment

This manuscript examines the use of a machine-learning based emulator of surface melt over the Greenland ice sheet, trained on output from a regional climate model and its firm model.

A: Thank you for the constructive feedback. We acknowledge the concerns and comments, and addressed them as described below.

The manuscript is relatively well written, although it does feel rather underwhelming – perhaps due to the focus on the development of an emulator / machine learning method, and this not being followed up by using the emulator to investigate a pressing science question. A similar concern would be around the novelty of the work not being especially clear. For example, the Introduction mentions using several machine-learning approaches, including neural-networks, yet never explains why the neural network approach used here was adopted, or how this work builds on existing work (including limitations). This would be my major concern #1.

A: We substantially changed the introduction, including more related work, while at the same time describing more clearly where our work is similar or different to existing work. Furthermore, we also include a short explanation on why we use neural networks (lines 45-52 in the revised version).

I also feel the language could be a lot tighter in places, and also that there is occasionally some information missing that would make it much easier to read. There also seems to be something missing from the text to make it a 'Cryosphere' paper. For example, the Conclusions mentions 'diverse climatic regimes of Greenland' – but there has been little acknowledgement or explanation of this up to now, so it seems rather too little too late, and also no references mentioned. This would be major concern #2.

A: We revised our manuscript to position it more clearly as a cryosphere study. We thus expanded the related literature section, added a systematic input selection study, and expanded the discussion section.

Furthermore, we considered the reviewer comments regarding ML terminology, and clarified or reformulated parts, where terminology may be confusing for a cryosphere audience.

Additionally, I simply don't see section 3 as a 'Results and discussion' section – it came across as simply explaining the results, and no discussion of them. A 'discussion'

thoroughly interprets, analyses, and explains the significance of the study's findings in relation to the research question and existing literature. The Conclusion section actually contains some interesting discussion points, so I would suggest bolstering that instead, and also adding references.

A: We revised the discussion section significantly, adding more interpretation and references.,

I also have a few small points that I noticed on first reading:

- NN is defined twice in the Introduction,

A: fixed

- Introduction mentions XGBoost and Neural Networks, yet these may not be immediately familiar with a reader. A sentence explaining the basis for these approaches here may be useful.

A: We added some short explanation in lines 48-51 in the revised version.

- section 2.1 could be tightened, as for example it's not immediately clear that the atmospheric forcing mention (first line of the second paragraph) is for the firm model, and also its not clear what time resolution the sub-daily SMB outputs are, which are subsequently aggregated to daily values (third paragraph).

A: We now introduced the subsection 2.1.1 and made clear that this refers to the DMIHH model only. Furthermore, we clarified that the SMB outputs are saved as daily aggregates directly (lines 114/115), and not aggregated afterwards.

- section 2.1 and elsewhere use subheadings as bold text such as 'Data cleaning'. I have to say I really don't like this approach, and find it rather lazy.

A: We replaced the bold subheading with standard third-level subsection numbering.

- given the importance of the atmospheric forcing used, I think some additional information on the HIRHAM5 model would be useful, and also on the appropriateness of using ERAI to force it. For example there is no mention of the spatial resolution. Albedo is mentioned later in the manuscript, yet there is no mention here of how this is computed by HIRHAM5.

A: We added further descriptions and references in 2.1.1.

- in section 2.1 there is no mention of how the firm model is spun-up.

A: Similar as to the discussion above, the details of the spin-up procedure are not critical here; what matters is simply that the model has been spun up, ensuring that melt production is not affected by transients arising from the initialization or the long-term memory of the model (Langen et al., 2017). We added a brief explanation in lines 113/114 in the revised version.

- In section 2.1 it's not clear what 'symlog' means, or what the variable x is.

A: symlog means "symmetric logarithm transformation", and x is the placeholder for the turbulent heat flux variables to be transformed; we will reformulate this sentence (lines 137-139).

- In section 2.1, the period 1980-1990 is 11 years, not 10. It might also be worthwhile explaining the differences between the training, validation, and test periods – as this is not intuitively obvious to someone not familiar with machine-learning. Also, the manuscript should include some justification for selecting these periods, such as why only a single year (2016) is used for the test period.

A: We added further explanation, see reply to referee #2.

- Figure 1 caption typo. Schema -> Schematic.

A: fixed.

- Section 2.2, 'regressing the surface melt based on atmospheric variables' is rather vague. What are these variables? How are they chosen? Presumably these are the predictor values?

The choice of the 'atmospheric variables' is explained just in the paragraph above (lines 180-184 in the revised version).

- Is Eq. 1 missing an explanation of what N is?

A: Line 193 "daily input variables for $N+1$ days", so it's the number of preceding days used as input.

- Section 2.3, typo 'of the of'

A: fixed.

- The captions could do with more information. For example, the caption for Fig. 2 does not mention what the various labels are in the figure. SW, SE, CE etc. The caption for Table 2 does not mention what R^2 is.

A: We expanded the figure caption accordingly.

- Line 255 mentions 'residuals', but what this means is not defined – is it the difference between the actual and emulated values?

A: Residual means the predicted value minus the true value. We added the definition in line 309 in the revised version.

- Conclusion section uses 'neural network' and not NN.

A: fixed.

- For the Conclusion section, I would recommend adding mentioning the appropriate figure or table when the results are being reiterated, so the reader is absolutely clear about the novelty of the work.

A: [We added references to the associated figures and tables.](#)