

## 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 firn model.

A: Thank you for the constructive feedback. We acknowledge the concerns and comments, and will adapt our manuscript accordingly as described in our responses 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 agree that the novelty must be stated more explicitly in the introduction. Existing related studies focus on regressing monthly or annual firn properties, where day-to-day variability is not present. In contrast, our work targets daily surface melt, which exhibits high temporal variability compared to monthly/annual data, and slower evolving data like temperature within the firn pack. In the revised version we will expand the introduction to include a more comprehensive review of related work, and discuss more thoroughly how our work differs from existing work, and where it builds on existing work. To our knowledge, a similar emulator for this application has not been published yet, although the interest in the development of ML-emulators in cryosphere research is growing.

Furthermore, we plan to include a new systematic feature removal analysis on input variables in the revised version, which we see as critical and very insightful for the further development of ML-based emulators for SMB processes. We believe this addition will strengthen the scientific positioning of our work and may serve as a useful reference for future methodological development in this research direction.

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 acknowledge the reviewer's concern that the manuscript should more clearly position itself as a cryosphere study. Our work demonstrates the application of ML methods to SMB modelling. While the application of ML within cryosphere science is still developing, its use has expanded rapidly in recent years. We therefore see this study as illustrating how these methods can complement and enhance existing modelling approaches used in cryosphere research, rather than being primarily a contribution confined to ML literature. We are confident that an expanded related literature section, the added systematic feature removal analysis, and an in-depth discussion will address this gap.

We also acknowledge that the language can be improved, and will revise the manuscript throughout to improve clarity, taking into account the line comments. We will add additional context to improve readability, considering the comments from all reviewers; also we will deepen the discussion and add further references, including these relating to the diverse climatic regimes (Fettweis et al., 2020; Lenaerts et al., 2020; The IMBIE Team, 2020; Vandecrux et al., 2019; Wang et al., 2021).

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 agree that section 3 is currently more focused on presenting the results, rather than discussing and interpreting the findings. We will thoroughly revise this section and will expand on the discussion points mentioned in the conclusion to improve the discussion. The inclusion of a new systematic feature removal analysis on input variables will also provide further material for scientific analysis and discussion, which will help strengthen the overall interpretation of the results.

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

- NN is defined twice in the Introduction,

A: We will fix this.

- 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 will shortly introduce the mentioned approaches and add appropriate references.

- 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 see that this section causes confusion, and we will first rename the section to make it clear that this part refers to the firm model DMIHH only. We will also revise the section itself, and clarify that the SMB outputs are not aggregated in an additional post-processing step, but are saved as daily values directly, while the internal time stepping of the model is hourly.

- 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 will replace the bold subheading with standard third-level subsection numbering to ensure coherent formatting across all section levels.

- 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: The manuscript already includes a description of the forcing dataset, but we agree that several aspects can be clarified further for readers. The primary objective of this study is the development of the emulator based on a dataset with internal consistency between the atmospheric forcing and the resulting surface melt production. The assessment of HIRHAM5 under ERA-Interim forcing is not relevant for this study, and we see it more as a distraction than useful information for this work. However, the HIRHAM5 model forced by ERA-Interim has been used and validated by various studies (Langen et al., 2017; Lucas-Picher et al., 2012; Mankoff et al., 2021), and we will add the additional references together with the spatial resolution of the used HIRHAM version in the revised manuscript.

Also, the internal calculation of albedo in the DMIHH model is mentioned in lines 82-84. Here again, the explicit computation of albedo is not relevant; but the key point is that albedo is calculated based on the atmospheric variables, and is therefore consistent with the energy balance, and inherently included in the atmospheric data. The importance and the role of albedo for the ML-emulator is stated in line 143, where all the emulator inputs are motivated. However, we will revise this parts to make this information more explicit to the reader.

- 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 will add a brief explanation of the spin-up and its importance to the revised manuscript.

- in section 2.1 its not clear what 'symlog' means, or what the variable x is.

A: symlog means "**sym**metric **log**arithm transformation", and x is the placeholder for the turbulent heat flux variables to be transformed; we will rephrase this sentence for clarity.

- 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 will add some explanation on train/val/test splits of sequential data in general, and selection of subset sizes, backed up by some references (Auffarth, 2021; Géron, 2019). Furthermore, we will add a justification for selecting 2016 as our only test year (see reply to referee #2).

- Figure 1 caption typo. Schema -> Schematic.

A: We will fix this typo.

- 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 138/139), i.e. air temperature, rainfall, snowfall, the turbulent heat fluxes, and the downwelling radiation. We refer to them as atmospheric variables, because "predictor values" (or also "predictor variables", "inputs", "input features"), typically refers to the specific preprocessed inputs of a model (which are denoted as  $X_d$  and  $X_i$ ; lines 140-142).

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

A: Line 155 "daily input variables for N+1 days", so its the number of preceding days used as input. We will rephrase this part to make it clearer to the reader.

- Section 2.3, typo 'of the of'

A: We will fix this typo.

- 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 will expand the figure and table captions to provide more complete information so that they can serve as more stand-alone sources of explanation.

- 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 will define this explicitly in the revised version.

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

A: We will fix this.

- 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 will rework the conclusion to make it easier to read.

## References

- Auffarth, B. (2021). Machine learning for time-series with Python. Packt Publishing United Kingdom.
- Fettweis, X., Hofer, S., Krebs-Kanzow, U., Amory, C., Aoki, T., Berends, C. J., Born, A., Box, J. E., Delhasse, A., Fujita, K., Gierz, P., Goelzer, H., Hanna, E., Hashimoto, A., Huybrechts, P., Kapsch, M.-L., King, M. D., Kittel, C., Lang, C., ... Zolles, T. (2020). GrSMBMIP: intercomparison of the modelled 1980–2012 surface mass balance over the Greenland Ice Sheet. *The Cryosphere*, 14(11), 3935–3958. <https://doi.org/10.5194/tc-14-3935-2020>
- Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media, Inc.
- Langen, P. L., Fausto, R. S., Vandecrux, B., Mottram, R. H., & Box, J. E. (2017). Liquid Water Flow and Retention on the Greenland Ice Sheet in the Regional Climate Model HIRHAM5: Local and Large-Scale Impacts. *Frontiers in Earth Science*, 4. <https://www.frontiersin.org/articles/10.3389/feart.2016.00110>
- Lenaerts, J., Camron, M. D., Wyburn-Powell, C. R., & Kay, J. E. (2020). Present-day and future Greenland Ice Sheet precipitation frequency from CloudSat observations and the Community Earth System Model. *The Cryosphere*, 14(7), 2253–2265.
- Lucas-Picher, P., Wulff-Nielsen, M., Christensen, J. H., Aðalgeirsdóttir, G., Mottram, R., & Simonsen, S. B. (2012). Very high resolution regional climate model simulations over Greenland: Identifying added value. *Journal of Geophysical Research: Atmospheres*, 117(D2). <https://doi.org/10.1029/2011JD016267>

- Mankoff, K. D., Fettweis, X., Langen, P. L., Stendel, M., Kjeldsen, K. K., Karlsson, N. B., Noël, B., Van Den Broeke, M. R., Solgaard, A., & Colgan, W. (2021). Greenland ice sheet mass balance from 1840 through next week. *Earth System Science Data*, 13(10), 5001–5025. <https://doi.org/10.5194/essd-13-5001-2021>
- The IMBIE Team. (2020). Mass balance of the Greenland Ice Sheet from 1992 to 2018. *Nature*, 579(7798), 233–239. <https://doi.org/10.1038/s41586-019-1855-2>
- Vandecrux, B., MacFerrin, M., Machguth, H., Colgan, W. T., van As, D., Heilig, A., Stevens, C. M., Charalampidis, C., Fausto, R. S., Morris, E. M., Mosley-Thompson, E., Koenig, L., Montgomery, L. N., Miège, C., Simonsen, S. B., Ingeman-Nielsen, T., & Box, J. E. (2019). Firn data compilation reveals widespread decrease of firn air content in western Greenland. *The Cryosphere*, 13(3), 845–859. <https://doi.org/10.5194/tc-13-845-2019>
- Wang, W., Zender, C. S., van As, D., Fausto, R. S., & Laffin, M. K. (2021). Greenland Surface Melt Dominated by Solar and Sensible Heating. *Geophysical Research Letters*, 48(7), e2020GL090653. <https://doi.org/10.1029/2020GL090653>