Review of ”Physics-aware Machine Learning for Glacier Ice Thickness Estimation: A case Study for Svalbard” by Steidl and al.

This paper proposes a framework based on Physics-aware machine learning to estimate the ice thickness distribution of grounded glaciers, applied to seven glaciers located in Svalbard. To achieve this, the authors incorporate mass conservation and other physical constraints into the cost function to be minimized during the training stage. They use ice thickness change and surface mass balance as primary constraints for the flux divergence. The authors conduct a glacier-wise cross-validation and discuss their results, and the cause of the rather-disappointing generalization in view to broader application.

This paper aligns with a series of recent works that explore the significant potential of machine learning techniques for ice thickness inference [e.g. Haq et al., 2021, Teisberg et al., 2021, Jouvet, 2023], including physics-informed deep learning [e.g. Riel and Minchew, 2023, 2022, Iwasaki and Lai, 2023, Jouvet and Cordonnier, 2023]. Among these, Physics-Informed Neural Networks (PINN) [Raissi et al., 2019] are particularly promising and represent an active research area.

Overall I find the approach interesting and I acknowledge the authors for their effort in presenting their results honestly. However, the work appears to be a work-in-progress, and lacks a clear picture of what should be done for making the method generalizable. This is partly due to a too a quickly-made and not very smooth writting (see e.g. my comments below about Section 2.3). The possible causes for the poor generalization results (LOGO CV) should be investigated (especially their relative importance) rather than listed, and I feel that this work could yield improvement in generalization already in this manuscript. I have a series of recommendations (encompassing the writing the paper and the methodology) that I hope will help the authors to improve the manuscript, which has the potential for a larger impact.

- In general, I find the description of the PINN rather inefficient. I focus here on Section 2.3, but my comments may be extrapolated to the entire paper. Do not expect TC readers to be familiar with all ML machinery, even less so with Physics-Informed ones. Therefore, you should provide a minimal background. Currently, section 2.3 is addressed solely to people with prior knowledge. At a minimum, briefly explain what a neural network is (a sequential composition of linear and nonlinear functions with optimizable weights). To my knowledge, this term can be intimidating, while it is not that complicated provided a minimal explanation. The current Section 2.3 mixes crucial information (I/O of the PINN) with more technical details (e..g activation functions, which are important but not essential for most readers unfamiliar with ML). I suggest distinguishing these two levels when rewriting this part to smooth the reading and target a broader audience. Consider moving ML technicalities to an appendix, and leave te ideas in the body of the paper. Another example: you mention ”Fourier layers” but do not provide any rationale (I would like to know the benefit of this). There are several ML-specific concepts (e.g., unlabeled data) that are not explained throughout the manuscript, which is a problem to maximize the audience of the paper to a general glaciological audience.

- I am not sure I understand: Do you feed your neural network with raster data grids (as suggested in Fig 1) or with large vectors of data at each coordinate along with the coordinate data? My question is whether you exploit the spatial structure of the data (I assume you have data on a raster structure grid). If not, I understand why you use a fully conneced network; if you do, why not use a convolutional neural network designed to capture spatial relationships?

- The description of ice flow (Section 2.2) seems rather simplified. There are a couple of assumptions behind that are not clearly written down. Including a true high-order model here would be a great added value I think. You mention in line 274 that adding momentum conservation would be "technically easy," but I am less pessimistic than you about the claim that" it would complicate the optimization of the model." Instead, the functional associated with the Blatter-Pattyn model, for example, behaves relatively well with good convexity properties [Jouvet, 2016, Jouvet and Cordonnier, 2023], and could act as a physically-consistent, welcome smoother.

- In connection with my previous point, have you considered moving the surface velocity from the input of your PINN to the data? This would make sense if you are including momentum conservation. In the present case, can this be an option too? What is the motivation to insert the "observational" data in input of the PINN or as data constrained in the loss?

- The comparison (Section 4.2) to the two other products [Millan et al., 2022, Farinotti et al., 2019] is not a strong point. It tells us that the PINN lies within the range, which is not surprising as the two products differ significantly. This section could be moved to an appendix.

- Maybe consider applying your method first to a synthetic case where you can create a manufactured bedrock and dataset. Then, use your method to infer the ice thickness and compare it to the ground truth. This approach would help avoid issues related to data suspicion. In general, there are many possible causes for the lack of generalization, but there are strategies to isolate these causes that you could further explore through synthetic experiments.

- Section 5 provides a list of possible causes for the lack of generalization. However, it is hard to draw any conclusions. Some causes are more important than others. It would be helpful if you could prioritize these causes (and improvement items) by order of importance, from the most significant (with the largest potential for improvement) to the least significant. I feel that "Physical constraints" should be at the top of the list.

- Lines 264-266: You place a lot of trust in your Mass balance reconstruction, especially if it is not calibrated (line 264). Considering that this is a major constraint, I think this might be a significant cause of underperformance. Also, using a model for estimating the SMB (even a perfected one) is problematic, as your "observations" are not observations but modelled reconstructions. Have you considered using in-situ sparse measurements instead?

I have some additional specific comments:

- In the introduction, it would be good to ellaborate the existing literature on using ML for ice thickness inversion modeling [e.g. Haq et al., 2021, Teisberg et al., 2021, Jouvet, 2023] (line 21), as well as physics-informed deep learning applied to similar problems, such as inferring basal conditions (bedrock location or slipperiness) [e.g. Riel and Minchew, 2023, 2022, Iwasaki and Lai, 2023, Jouvet and Cordonnier, 2023] (lines 32-34). As this is a fast-evolving field, it would be good to check the latest papers, and possibly to complete.

- l 12: Not sure Millan et al. [2022] is the most appropriate reference for that.

- l 13: "Physics-based approaches ..." This sounds to be a very personal definition, consider a more appropriate one.

- l 22: "One avantage of data-driven approaches is a significant speed-up compared to physics-based models" : The computation speed-up has nothing to do with whether it is data-driven or physics-driven; it is the result of the efficiency of evaluating a neural network (especially on GPUs), irrepective of the training strategy: based on data [Jouvet et al., 2022] or on physics [Jouvet and Cordonnier, 2023]. Please correct.

- l 30-31: These two sentences are unclear to me : i) what means "data-efficient" in the context? ii) "boundary condition to solve the PDE", I think I understand what you mean (this would be a Dirichlet BC as you can enforce the solution to be close to a certain given value somewhere), but I'm not sure this is clear for all.
- l 255: "the loss landscape is highly complex", this is an unusual way to describe the lack of convexity the loss, which is not improved – I agree – by adding the number of constraints within the loss. I am not sure I found what optimizer you used (ADAM, SGD, RMSPROP, ?).
- Appendix B: I feel I have seen this exercise numerous times in textbooks, deriving a 0.8 ratio between vertically-averaged and surface velocity in the non-sliding SIA parallel slab case. I suggest you replace it a reference and use the space in the paper to better explain the ML part.

References

D. Farinotti, M. Huss, J. J. Fürst, J. Landmann, H. Machguth, F. Maussion, and A. Pandit. A consensus estimate for the ice thickness distribution of all glaciers on earth. *Nature Geoscience*, 12(3):168–173, 2019.

M. A. Haq, M. F. Azam, and C. Vincent. Efficiency of artificial neural networks for glacier ice-thickness estimation: A case study in western himalaya, india. *Journal of Glaciology*, 67(264): 671–684, 2021.

Y. Iwasaki and C.-Y. Lai. One-dimensional ice shelf hardness inversion: Clustering behavior and collocation resampling in physics-informed neural networks. *Journal of Computational Physics*, 492:112435, 2023.

G. Jouvet. Mechanical error estimators for shallow ice flow models. *Journal of Fluid Mechanics*, 807:40–61, 2016.

G. Jouvet. Inversion of a stokes glacier flow model emulated by deep learning. *Journal of Glaciology*, 69(273):13–26, 2023.

G. Jouvet and G. Cordonnier. Ice-flow model emulator based on physics-informed deep learning. *Journal of Glaciology*, pages 1–15, 2023.

G. Jouvet, G. Cordonnier, B. Kim, M. Lüthi, A. Vieli, and A. Aschwanden. Deep learning speeds up ice flow modelling by several orders of magnitude. *Journal of Glaciology*, 68(270):651–664, 2022.

R. Millan, J. Mouginot, A. Rabatel, and M. Morlighem. Ice velocity and thickness of the world's glaciers. *Nature Geoscience*, 15(2):124–129, 2022.

M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

B. Riel and B. Minchew. High-dimensional flow law parameter calibration and uncertainty quantification over antarctic ice shelves: a variational bayesian approach using deep learning. *Authorea Preprints*, 2022.

B. Riel and B. Minchew. Variational inference of ice shelf rheology with physics-informed machine learning. *Journal of Glaciology*, 69(277):1167–1186, 2023.

T. O. Teisberg, D. M. Schroeder, and E. J. MacKie. A machine learning approach to mass-conserving ice thickness interpolation. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 8664–8667. IEEE, 2021.