This work applies a Physics-Informed Neural Network (PINN) as a data-driven tool to estimate ice thickness across glaciers located on Spitsbergen in Svalbard. A physics-based loss function used in training the PINN is designed to penalize solutions diverging from a modified form of mass conservation. Additional physics-inspired loss functions are added relating components of the surface velocity arising from deformation of the ice and sliding at the base. The neural network is also provided with a number of other data inputs, including surface velocity, surface slope, elevation, positional parameters, and values providing assumed relationships between surface and depth-averaged velocities. The authors explore their results using a cross-validation scheme designed to avoid problems with spatial correlation.

PINNs have seen an increasing number of uses within glaciology in the past few years. Their intrinsic ability to mix together known physics with poorly calibrated constants and sparse and/or noisy measurements make them an appealing modeling tool for underdetermined problems. This work is novel in training a single PINN over an extremely large domain (effectively all of Spitsbergen) and in mixing a large number of physical constraints, physically-inspired constraints, and plausibly related data sources.

Two related challenges complicate the evaluation of results in this work. First, as with almost all ice-covered regions, direct measurements of ice thickness are sparse. Second, the dynamics of the glaciers are complex and poorly understood. In Svalbard, a number of complicating factors are at play:

1. Many glaciers are topographically constrained, making lateral drag important and complicating simplified models of ice dynamics.
2. Many glaciers are thought to be polythermal, often with a significant layer of temperate ice at the base overlain by cold ice (Sevestre et al., 2015)
3. Cold surface temperatures allow for the accumulation of thick firn layers with poorly constrained density (Pälli et al., 2017)

These complications are not unique to Svalbard, of course, but aspects of Svalbard's topography and geographic location make them especially notable here.

The authors frame the cross validation results in a way that seems somewhat disappointing. I am perhaps more optimistic than the authors about the results. In particular, I think the evaluation of a physically-based model on a glacier where no ice thickness data was provided is an unfair assessment of the model. The PINN proposed in this work is something of a hybrid between a data-driven estimator and a PDE solver. These two types of tools would be accessed in different ways. Additional consideration of appropriate evaluation mechanisms is probably needed.

Architecturally, I think this work is very interesting. There is a novel fusion of physics-based, physics-inspired, and non-physical relationships at work here. Unfortunately, the lack of explainability and the lack of a good ground-truth data source make it difficult to see a path to the results presented here significantly updating our thinking about Svalbard's glaciers.

Given this combination, I would encourage the authors to consider leaning into exploring the design of the PINN by, for example, exploring the importance of the various input fields or designing an experiment to consider the use of different ice physics approximations within this framework.

Specific comments below:

PINN

- It is not clear to me what the coordinate system is used to feed the network. Is it a standard projection? Are the coordinates consistent across all of Spitsbergen or are glaciers each on their own local coordinate system in some way?
- Why the current set of inputs to the neural network? It is not obvious to me, for example, why the area of the glacier should be included. In general, it would be interesting to know how including each input impacts the results.

Physical Model

- The way that deformation velocity, sliding velocity, depth-averaged velocity, and surface velocity are explained is somewhat confusing to me. My interpretation is that the authors are using a simplified physical model (Appendix B) to set a relationship between surface and depth averaged velocity, which you then qualitatively decide to loosen. Separately, they assume that sliding velocity and surface velocity are related by a pre-determined field. The network predicts deformation velocity only and evaluation of the mass conservation loss term is done by adding in sliding velocity according to the defined constant and the surface velocity.
    - What is the significance of the network outputting deformation velocity rather than directly depth-averaged velocity? Lines 61-62 seem to imply this is important, however it is not clear to me why. It seems to me that it is simply a choice between an extra calculation to compute mass balance and an extra calculation to compute the depth-averaged velocity bounds loss.
    - Lines 69-70 state that depth-averaged velocities are calculated for the x direction, y direction, and magnitude separately using different values of beta. Beta relates surface velocity to sliding velocity. In the simplified model of Appendix B, the sliding velocity must be in the same direction as the surface velocity, but different values of beta for x and y implies that the sliding velocity is in a different direction.
- Apart from stating that ice is assumed to be incompressible (Line 50), I saw no mention of the effects of unknown density of snow and firn. To my understanding, glaciers in Svalbard may have significant firn layers (Pälli et al., 2017). This contributes to uncertainty in the radar measurements (as the dielectric permittivity is dependent on density) and impacts the implied mass flux. This source of uncertainty should at least be discussed.
- In my view, the simplified ice dynamics of Appendix B may be insufficient for glaciers in Svalbard. I believe that the model selected ignores stresses from drag against the

sidewalls, which seem significant for the topographically constrained glaciers on Svalbard. Additionally, assuming A to be constant with depth seems like a stretch. Many glaciers are suspected to be polythermal and this has been proposed as a mechanism for the surge behavior seen in Svalbard (Sevestre et al., 2015). While the authors have excluded currently surging glaciers, the presence of this phenomenon implies to me that depth-dependent temperature may be an important part of glacier dynamics in this region. At a minimum, further discussion of this point is needed.

OGGM-Processed Inputs

- Are any of the input fields that are processed with OGGM interpolated by OGGM in any way? If they are interpolated following a similar physical model to yours, does this introduce a circularity?
- I think it would be helpful to discuss how the surface mass balance input is derived. It sounds like a model-derived value? There are quite a few weather stations in Svalbard. Has the model been validated? How does it perform?

Training and Evaluation

- The authors point out that data is highly correlated in space and thus they have used a cross-validation scheme based on leaving out an entire glacier at a time. I think that's a good approach to a challenging issue.
- With the above said, however, I do wonder if this is an overly harsh method of evaluation. The effect is that, in looking at Table 2, we're looking at glaciers where no ice thickness data was available, greatly diminishing the value of the mass conservation approach. Another approach might be to leave in only the highest (elevation) 20% of the ice thickness data and explore how well the PINN can use mass conservation to extrapolate this downstream.
- On Line 201, the authors state that the results suggest the model is overfitting. While this would be the conventional interpretation for a neural network, I think this is an overly critical interpretation for a PINN. Evaluating a PINN with no training data for the data loss function is sort of like evaluating a PDE solver with no boundary conditions. The analogy does not fully hold as the authors have also introduced some other inputs which can perhaps be used to guess at the ice thickness, but, in general, I think the authors may be too critical of their own results here.
- Later (Line 177), the authors mention a random split between training and validation data. Given the aforementioned spatial correlation problem, how is this validation dataset used? Is it meaningfully independent of the training data?

Interpretation and Applications

- It would be good to discuss the importance of ice thickness on Svalbard. This might depend on what you think your model is good at. For example, an improved estimate of total ice volume would be impactful for sea level rise projects. Improved fine-scale ice-free topography might have more relevance to projecting the evolution of specific glaciers that are relevant to local communities.

- I would like to see discussion of what components of the inputs and loss function are most important. Many applications of PINNs largely use them as tools for solving PDEs where constraints, regularizations, or boundary conditions do not easily fit in conventional solvers. This work goes beyond that, feeding in multiple layers of data that is not directly incorporated into a physics-based loss term. This, of course, raises the question of which parts are most informative. A careful set of experiments exploring this would be very interesting.

Typos and minor corrections

- Line 10-11 - Ice flux is determined by more than simply ice thickness and surface slope under real world conditions. This should be clarified to not suggest that those two variables alone are sufficient.
- Line 69 - bracket is the wrong way around
- Figure 4 - Are the color scales saturating? If so, it would be good to show the clipping in a different color so we can see where the error exceeds +/- 100 m.
- In Table 2, comparing the first glacier’s performance in-sample versus LOGO, the RMSD more than doubles while the MAPD decreases. Is this correct?

I enjoyed reading this work and believe it to be a promising avenue. I hope that these comments can help improve this manuscript.