Response to the Reviewers

January 4, 2026

Response to Reviewer 2

This paper by Krishna et al. introduces a neural-network-based approach to infer bed topography constrained by both conservation of mass and momentum. The authors find that this approach can infer more physically realistic bed topography in slower-moving regions than BedMachine, which is only mass-conserving. The paper is well written and presents a valuable comparison between the two approaches. As PINN is a new method, such methodological studies are important for building an understanding of their benefits and current limitations. I have a few comments, summarized below:
Response: We appreciate the reviewer's thoughtful comments. Thank you for your positive feedback and for highlighting the relevance of our methodological study.

First, it's a good idea to hide part of the thickness data for independent validation in Appendix C. As $b = s - H$, the error reporting in the paper for b is a direct result of the errors in s and H, for which you have training data. My understanding is that the error reported in the main text between the prediction and the "ground truth" topography, along locations where H and s are provided to the NN, simply reflects how well the NN fits the H and s data. It is not relevant to the quality of the physics-based interpolation. A small error along the flight tracks does not imply good b prediction between the tracks, so I'd be cautious about mixing the interpretation of data error in b with the success of the interpolation. That said, in Appendix C the prediction of b at locations without H data is the true demonstration of how effective the physics-informed interpolation is in regions where data are not directly available.
Response: Thank you for your comment, and for your positive feedback with respect to the validation study in Appendix C. Regarding the point about the uncertainty in the training data, we have indeed assumed that this data is the "ground-truth". We have tried to be careful about reporting the differences between predictions and the training data as "RMSEs" and have tried not to use the terms "error" or "uncertainty". However, to minimize confusion, we will explicitly state in table captions that the values reported are "differences" or "RMSEs" that should not be confused with "error" or "uncertainty". We will also explicitly discuss how the errors in the training data might affect the quality of our PINN predictions in the discussion section.

Second, the prediction of velocity from SSA is the depth-averaged velocity, but the observed velocity in the loss function is the surface velocity. Can the authors comment on when this distinction is important and why it is justifiable in the paper?Additionally, is SSA a good approximation of Stokes in all regions studied by the authors?
Response: It is indeed true that depth-averaged velocities are not always the same as surface velocities, which are used in the PINN loss function. Morlighem et al., (2011) which is one of the first references using a mass-conservation based approach for inferring glacial thickness accounts for this by incorporating a factor of 0.95 and a tolerance interval of ± 50 m/yr for ice velocity within the optimization process (assuming that the depth-averaged velocities can be be 5 percent smaller than the surface velocity). Moreover, while the PINN loss function is set up to minimize differences between predicted velocities and the surface velocities, it also minimizes the mass balance and momentum balance residuals. As a result, the predicted velocity fields are unlikely to be identical to the surface velocity fields. Indeed, we have shown in Table 4 that the RMSEs of the velocity fields for each of the three regions are on the same order of magnitude as the tolerance interval in Morlighem et al. (2011), and consequently in BedMachine.
With regards to the use of SSA, several modeling studies, including models in the ISMIP6 ensemble (Goelzer

et al., 2020) use SSA to estimate the future sea level contribution of the Greenland Ice Sheet. That being said, it is also important to note that, in this case, we decided to start with a simpler approximation of the momentum balance to minimize computational cost and to see if this method of using two conservation laws could yield sensible results. Moreover, since we wanted our method to be comparable to BedMachine, we needed to use a 2D approximation of the momentum balance. For future work and for applying this method for the interior, slower-moving region of the ice sheet, we hope to use higher-order approximations of the momentum balance. We will elaborate on this in our discussion section.

Third, why does the mass-conserving PINN (MS) produce results so different from BedMachine (Fig. 4), if they both conserve mass? It is mentioned that PINN (MS) predicts isolated, unrealistic crater-like features along radar flight tracks due to "overfitting" (line 323), but why does the mass-conserving BedMachine not suffer from the same issue in the same region? It appears that PINN is using less data than BedMachine, but are they using the same thickness data? Can this difference in input data between PINN (MS) and BedMachine be made more explicit?
Response: The PINN (MB) and PINN (SB) do predict isolated bed features along the radar-flight tracks. However, these isolated features are located in the region *outside* of the mass-conserving domain of BedMachine, $\Omega \setminus \Omega_{\mathrm{mc}}$, as stated in the manuscript. In this area outside the mass-conserving domain, BedMachine uses other interpolation methods like kriging, whereas the PINN is attempting to constrain predictions to the conservation of mass and the conservation of momentum respectively. We have described this in our discussion section, but will work to make it clearer to minimize confusion.
With regards to the difference between the PINN training process and the BedMachine optimization process, we have mentioned that BedMachine uses all of the ice thickness training data whereas the PINN approach involves randomly selecting data points within the domain of interest in Appendix C. We have cited the BedMachine paper (Morlighem et al., 2017) where the details are well explained and documented, however, we will make this difference more explicit in our discussion.

Fourth, As PINN solves the equations weakly, the only measure of success, apart from data misfit, is the equation residual. In addition to the various data errors, I believe the paper needs to show map views of the equation residual to demonstrate convincing training success, and the PDE residual should be evaluated on a higher density of collocation points than the training collocation points to check whether the PDE is satisfied between the training collocation points.
Response: Thank you for the feedback. We will include these additional figures of the mass balance and momentum balance residuals in our manuscript.

Fifth, in the comparison of errors between different inversion results, it is only meaningful to say A is lower than B if the difference between them is larger than the uncertainties in A's errors. There are many discussions of error comparisons between PINNs and BedMachine, but the PINN errors are averaged over an ensemble of PINN predictions. It is important to consider the spread of errors among PINN predictions. In error reporting such as Table 2, I highly recommend including not only the error of the mean PINN prediction, but also error bars representing the range of error for each PINN prediction. Comparisons between errors are only meaningful after including the uncertainties in PINN errors due to the ensemble.
Response: We will also include figures that depict the standard error / range of the PINN ensemble for each region in our manuscript.

Finally, would it be possible for the authors to comment on the feasibility of enforcing both momentum and mass balance in the classical adjoint method? If this would be difficult, that would also strengthen the paper's narrative. This is a natural question that readers are likely to have and will be eager to hear the authors address.
Response: Using the classical adjoint method for this problem (i.e., in this case, attempting to invert for ice thickness and basal sliding coefficient using both conservation of mass and conservation of momentum) is technically difficult. The derivation with traditional adjoint method for a flow-line SSA problem can be found in Cheng et al. (2021), however, numerical implementation of 2D problem will require automatic differentiation. We will add this in the discussion to strengthen the paper's narrative.

Minor comments

Line 57-59: Literature review: Bolibar et al., 2023 did not use PINN. It is correct that Riel et al. (2021) is the first PINN study in glaciology. But Riel and Minchew (2023) appeared after some of the other cited studies, and thus I recommend moving it into the sentence along with other papers.
Response: We will make these changes in the manuscript.

Line 85: "satisfy the PDE residuals" → "satisfy the PDEs"
Response: Thank you for the suggestion, we will make this change in the manuscript.

Figure 1: Nice figure. Subscript "data" is missing in $L = L_{data} + L_{\phi}$
Response: Thank you for catching this typo! We will re-make the figure accordingly.

Line 86-87: Are your collocation points fixed in location or changing throughout iterations? Changing throughout iteration is highly recommended; if not doing so you'll likely overfit the physics on the fixed discrete collocation points.
Response: While it is possible to re-sample the collocation points during the training process and is a feature built-in to the PINNICLE package (Cheng et al., 2025), it can add to the computational cost because resampling often leads to slower convergence. We have chosen a high number of collocation points for this reason, so that the physical laws are generally conserved throughout the domain. Moreover, we also trained an ensemble of 5 PINN models with different random seeds. This resulted in the random selection of different collocation points for each of the PINNs, thereby reducing the likelihood of the PINN satisfying the physics for fixed collocation points. For each location, we then took the median prediction of the 5 PINNs. We will make this clear in the methods section. We will add a new test that allows for re-sampling of the collocation points for one of the regions. Based on our previous experiences during the development of PINNICLE (Cheng et al., 2025), we expect this result to be similar to the results of the experiments without resampling.

Line 115-117: As the apparent mass balance contains both ice thinning rate and the surface/basal mass balance, can you explicitly say how RACMO and ICESat-2 data are combined to give $\dot{a}$? Does ICESat-2 give a thinning rate $\partial H/\partial t$ without the effect of $\dot{M}_s, b$?
Response: Thank you for your feedback. We can make it more clear in the methods section that the surface mass balance $\dot{M}_s$ is from RACMO, the basal mass balance $\dot{M}_b$ is assumed to be negligible, and the ice thinning rates are derived from ICESat-2.

Table 1: Can you use the same time units (either year or second) between the weight values and the variable values for comparisons?
Response: We have chosen to report the weights in this manner in order to be consistent with the previous publications Cheng et al., (2024, 2025) and the PINNICLE documentation. The typical values are reported in m y^{-1} as this is easily understood by members of the cryosphere community and the ice is moving too slowly for the velocities to be reported in m s^{-1}. The weight values are reported in the SI units as these are the real values used in the computations.

Line 167: Regarding "the PINN output variables will have different values for different regions of the GrIS", are you using the same weights across the three different regions where the velocities can be very different?
Response: The weight values are the same for all thee regions. This framework generally works since the mean velocities over these regions are generally comparable since they include both fast-flowing and slower-moving ice. Moreover, we also included a logarithmic terms in our loss function to account for the differences in the velocities across the regions. Deception and Narssap generally have slower velocities than Upernavik, and these logarithm loss terms help better captures the slower velocities. We will make this more clear in the methods section.

Line 172: I like the fact that different random seeds allow you to sample different solutions that can solve the ill-posed inverse problem. Given the ill-poseness, and the importance of training an ensemble of PINNs, could you elaborate on why 5 time is sufficient, and if you expect different medians if you can train more PINNs?

Response: While increasing the number of PINNs in our ensemble is always desirable, we decided that 5 PINNs would be suitable for each region given the computational cost of training each PINN. We find that since we have chosen a high number of iterations for training each PINN in the ensemble, the PINN predictions generally converge and hence an ensemble of 5 PINNs is suitable.

Eqn 12: Why do you not need data loss for surface elevation s in the loss?

Response: The mass balance equation does not include a surface elevation term, hence we do not need the PINN to predict surface elevation values. As a result, we do not need to include this term in the loss function.

Line 205: Regarding "PINN (SB) is exposed to the ice velocity data along the boundaries of the region of interest, thereby satisfying the stress balance boundary conditions" How many velocity data points do you have along the boundary? I thought the velocity data is 400 points sampled within the domain, meaning that along the boundary of the ROI the velocity data points would be sparse, not truly "satisfying the stress balance boundary conditions".

Response: Thanks for pointing this out. We will rephrase this sentence to make it more clear within the manuscript. We do randomly select 4000 data points for $\boldsymbol{v_x}$ and 4000 data points for $\boldsymbol{v_y}$, and these include both points within the domain of interest as well as on the boundaries of the domain. It's true that most of these points will likely fall within the domain as opposed to along the boundaries. Using PINNICLE, we did not explicitly separate boundary points from the interior of the domain.

Additionally, as you're also solving for H and s, can you comment on the boundary conditions involving s and H? As they are within spatial derivatives in the momentum equation they also theoretically require BCs just like velocities.

Response: Similarly, we randomly select 4000 surface elevation and ice thickness data points from the training data sets. The domain boundary of surface is included as the velocity, however, there is no explicit domain boundary conditions for the ice thickness, because we are only using data along ice-penetrating radar flight tracks which do not usually follow the domain boundary. We assume that the ice thickness predictions are constrained as long as we have some ice thickness data along each flow line (Morlighem et al., 2011).

Somewhere in the paper write down σ_{SSA} in stress balance in terms of velocities; this will make the discussions of the velocity boundary conditions more clear.

Response: Thank you for the suggestion, we will make this change in the manuscript.

Figure 4 d-f: Does positive denote higher Bedmachine or PINN topography?

Response: We will definitely make this more explicit in the Fig. 4 and Fig. 5 captions. We will state, explicitly, that the difference is $\hat{\boldsymbol{b}}_{\text{BedMachine}} - \hat{\boldsymbol{b}}$ for panels d-f.

Line 259: Regarding "the PINN (MB+SB) RMSE of 137 m is 260 lower than that of BedMachine (see Table 2)." I think you really need error bars for the PINN RMSE to compare which one is lower.

Response: Thank you for this feedback. We will include additional information on the uncertainties and error estimates in an updated version of this manuscript.

Line 278: Regarding "Lastly, the PINN predicts a 'disconnected' trough beneath the southern fork of Upernavik Isstrøm North, while BedMachine suggests that this trough is continuous." Do you also see disconnected troughs at each PINN result, prior to the averaging across PINN results? When taking the mean between different PINN results do you remove some features that were apparent in each individual PINN results?

Response: This is a good point that will likely become clear once we plot figures of the range/standard error of ensemble predictions for the bed topography in the Upernavik region. While it is possible that we are excluding some information by taking the median across PINN predictions, in almost all the experiments we have performed, the PINN has not been able to recover the full trough beneath the southern fork of Upernavik Isstrøm North. That being said, the 'disconnected' part of the trough is shaded dark green which is very close to a bed elevation of 0 m (Figure 5a), indicating that while the PINN does not predict the full depth of the trough (as shown in BedMachine), it is quite close. We will add this point to our discussion!

Table 4: Why is thickness RMSE not included?

Response: The thickness (or bed topography) RMSEs are included in Tables 2 and 3. Table 4 is only meant for the other state variables in the PDEs that we are not trying to infer, i.e., velocity, apparent mass balance, and surface elevation.