Response to the Reviewers

March 10, 2026

Response to the Editor

Dear Editor, thank you for your encouraging comments and for highlighting the relevance of our work. As suggested, we have added more detail to the methods section and further justified our modeling choices. We have also included maps of the mass balance and stress balance residuals in the Appendix. Per Reviewer #2's suggestion, we have improved our error estimates and validated our results against an independent ice thickness data set. We have addressed your comments in greater detail as follows:

Collocation Points

The dependence of your results on the sampling strategy of these points is not assessed. I think there should be a smarter way than random sampling. You could either use information in your input fields or at least guarantee a good coverage of the domain.
Response: Fig. 1 depicts the coverage of the data points for each PINN (initialized with different random seeds). Note that the random seeds are different since these are randomly generated to initalize each PINN. Therefore, the 4000 data points selected for each PINN within the ensemble (for each region) are likely to be different since all the random seeds are different, guaranteeing good coverage of the domain; in other words, the PINN ensemble is exposed to $\lesssim 20000$ data points. We also note that the higher number of collocation points ($N_{\varphi} = 9000$) randomly selected for each PINN within the ensemble, allows the physical laws to be conserved throughout the domain as the PINN ensemble is exposed to $\lesssim 45000$ collocation points. It is possible to resample the collocation points during the training process to improve coverage of the domain, and we have added a new section in Appendix C that details the effect of doing so. We find that the PINN trained with resampling of collocation points yields very similar results to that of the PINN ensemble.

Thickness Observations

I particularly appreciated Appendix C as you show how well your approach performs if less thickness data is available. For this, you remove a small portion of the available measurements. Yet I wonder how good your approach is when spatially transferred into regions without data. I therefore wonder if you could use one of your three regions and withhold all measurements within a complete drainage basin of one of the outlet glaciers. In this way, the reader would get an impression on the transferability and an understanding on how dependent this approach is on well distributed measurements.
Response: As suggested, we have run more validation experiments for the Upernavik region and included these results in Appendix B. Our previous validation experiment exposed the PINN to ice thickness data ~ 15 km apart and hid most of the data along Upernavik Isstrøm North. We run two more experiments, including an experiment where the PINN is exposed to ice thickness data ~ 30 km apart and an experiment where the PINN is only exposed to ice thickness data close to the inflow boundary (~ 60 km away from the outflow boundary). We find that while the results for a track spacing of ~ 30 km is good enough to capture the general bed characteristics of the Upernavik region, only providing data close to the inflow boundary does not lead to reasonable results. Hence, we find that the PINN needs more data along flow in order to constrain ice thickness predictions.

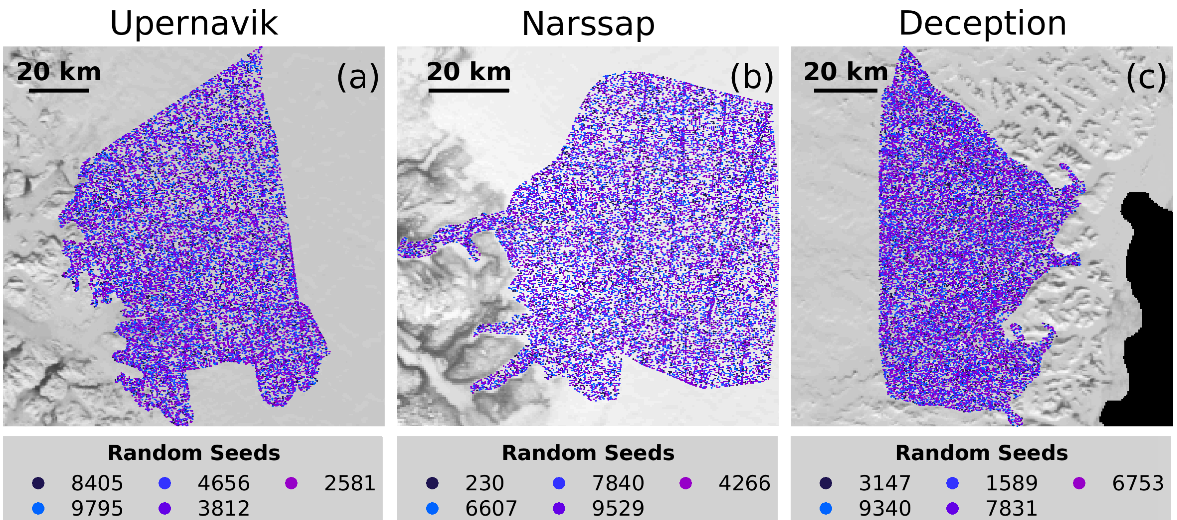

Figure 1: Data point coverage of the PINN (MB+SB) ensemble for (a) Upernavik, (b) Narssap, and (c) Deception. For each random seed, we plot the locations of 4000 data points (for some predicted variable).

Comparison to QRF Approach

The comparison to BedMachine forms a solid basis in your manuscript. It would be good if you specified the exact version of this product. I must have missed it.
Response: Thank you for catching that we didn't mention this! We were previously using a slightly older version of BedMachine, but in the recent revision we have updated all of the figures and our results of comparison with BedMachine to the latest version, BedMachine Greenland (Version 6). We have made this clear in the results section, all of the tables, and updated the citations appropriately.

On another note, the last author has recently contributed to a quantile regression forest for estimating the basal topography of Greenland (10.1017/jog.2025.10071). Together with the approach in the manuscript at hand, the reader might wonder about a comparison of these two recent products. I therefore consider a comparison as utile.
Response: As suggested, we have compared our bed topography results with the results generated by the quantile regression forest (Palmer et al., 2025), as shown in Fig. 2 below. We find that the QRF map for Upernavik is fairly similar to that of the PINN. However, the QRF approach predicts a discontious trough beneath Nunatakassaap Sermia and a discontinuous trough beneath the southern fork of Upernavik Isstrøm North. The QRF map for Deception is highly similar to the PINN result, as both capture the intricate network of troughs that coincide with the fast-flowing ice streams in the region. However, the QRF map for Narssap is quite different than that of the PINN, and it does not seem to detect the trough beneath Narssap Sermia. It should be noted that we have not included this comparison in our paper since the vast majority of modeling papers for the Greenland Ice Sheet use BedMachine Greenland, none use QRF (which was a proof of concept).

Moreover, there seems to be an independent dataset of thickness measurements for validation (PROMICE).
Response: Thank you for the suggestion. We have included a separate subsection within Appendix B, our updated validation section, comparing our PINN (MB+SB) ensemble median predictions with the independent ground-truth bed topography data set (PROMICE, Sandberg Sørensen et al., 2018) to strengthen our paper.

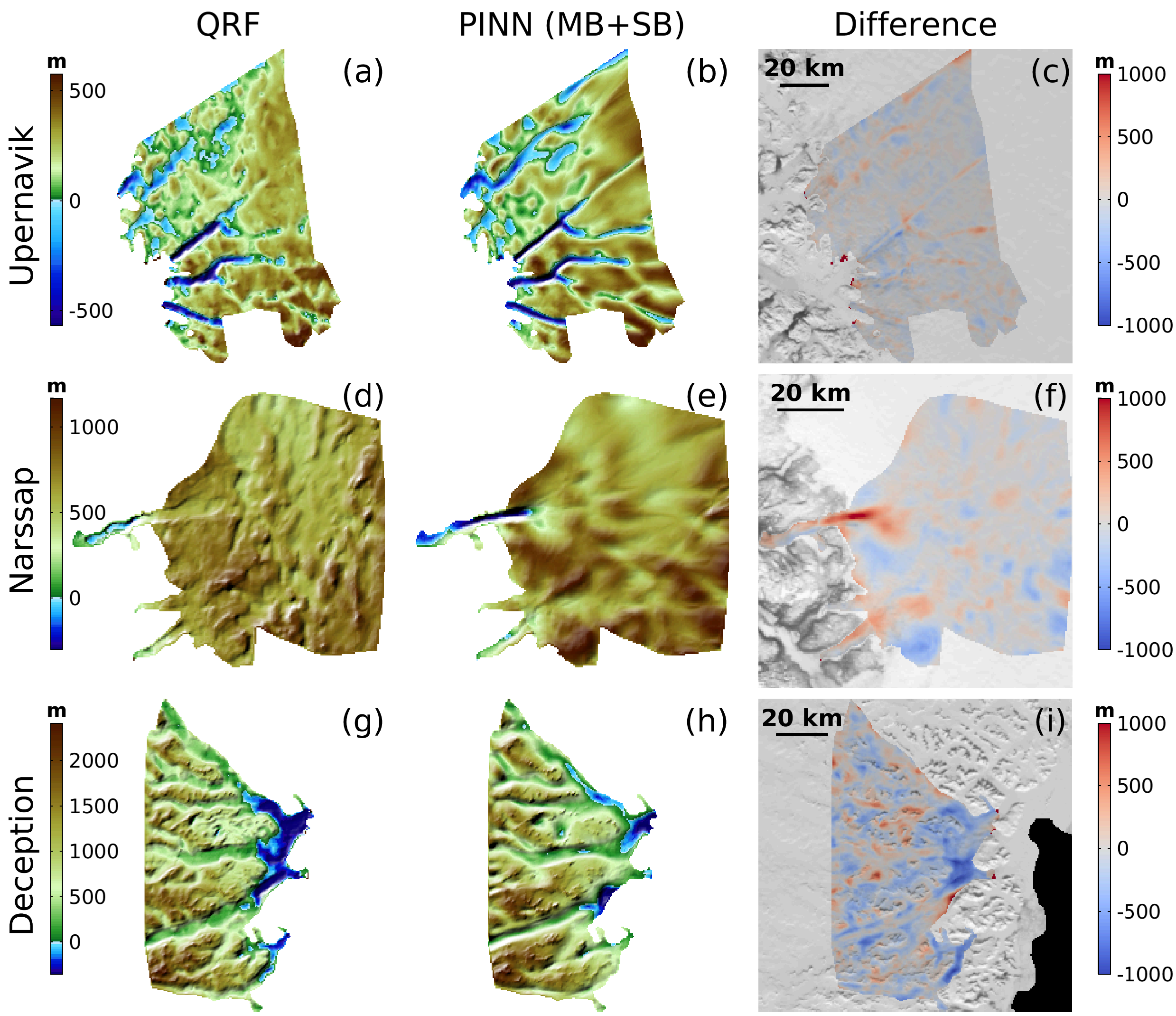

Figure 2: Comparison of PINN (MB+SB) predictions with Quantile Regression Forest (QRF) approach to estimating the bed topography (Palmer et al., 2025) for the (a,b,c) Upernavik, (d,e,f) Narssap, and (g,h,i) Deception regions. (a,d,g) depicts the QRF bed topography maps, (b,e,h) depicts the PINN (MB+SB) bed topography predictions, and (c,f,i) depicts the differences between the QRF and PINN (MB+SB) bed productions, $\hat{\boldsymbol{b}}_{\mathrm{QRF}} - \hat{\boldsymbol{b}}$ (m).

Response to Reviewer 1

Summary and general comments

This is a well-organized and well written paper describing the application of physics-informed neural networks (PINNS) for further improving the generation of sub-glacial bed topography datasets, building on previous "BedMachine" efforts that have been ongoing for the past decade or so. Two conservation equations – the continuity equation and a momentum balance equation for ice flow – are considered and introduced as additional constraints on the loss function, akin to their introduction as constraints in the "cost function" of PDE-constrained optimization (a good analogy to consider including for readers more familiar with the language of glaciology modeling / optimization?). The two constraints are considered on their own and in combination and the resulting bed topography datasets are compared and contrasted with previous BedMachine results for three glaciologically distinct regions. Overall, the authors argue convincingly that the new approach has merit and demonstrates potential for improving inferred bed topography in regions where the traditional BedMachine approach begins to break down. Overall, the work seems very worthy of publication and readers of The Cryosphere will find it a worthy contribution. My suggestion would be to accept for publication with minor revisions, noting that these are those suggested revisions (detailed below and identified by their line number in the submitted version) are largely editorial in nature.
Response: We appreciate the reviewer's thorough and thoughtful comments. Thank you for your encouraging feedback and for emphasizing the overall quality and relevance of our work.

My one more substantial suggestion – not necessarily for this publication but possibly for a future effort – is that I think it would be very useful to redo this exercise for a single region (I realize the computational cost could be a challenge, so pick a single region, like the most challenging one discussed here) but using L1L2 (Blatter/Pattyn) for the momentum balance model as opposed to SSA (which you've already done and could reuse as the baseline). Because the former allows for internal deformation, a possibly very different vertical velocity profile (and hence depth-averaged velocity and flux divergence) might be implied in regions of slower moving ice (noting that the modeled 2d surface velocity field could still be used as the velocity constraint, so that there should not necessarily be any significant reformulation of the loss functions discussed herein). It would be interesting to see if the more accurate stress balance constraint helped to alleviate any of the remaining problems discussed below.
Response: We appreciate the reviewer's feedback about using L1L2 (Blatter/Pattyn) for the momentum balance model to account for vertical shear. While this is currently outside the scope of this paper, we have included a note in the 'Limitations and Next Steps' section of the Discussion, emphasizing the need for more higher-order modeling especially in areas of slower-moving ice.

Detailed comments

21: Are these refs now considered the definitive source for defining the amount of potential SLR locked up in the ice sheets? If not, maybe consider adding one or two more from other authors for the sake of diversity?
Response: We believe that IMBIE Mass Balance Intercomparison (Otosaka et al., 2023) provides the latest consensus on the current contribution to sea level rise. We also believe that BedMachine Greenland (version 6) (Morlighem et al., 2017; Morlighem and et al, 2025), BedMachine Antarctica (version 4) (Morlighem et al., 2020; Morlighem, 2025) and Bedmap 3 (Pritchard et al., 2025) are the most up to date estimates for defining the amount of potential sea level rise. We have included these in our manuscript.

23: "numerical ice sheet modeling"
Response: Thank you for the suggestion, we have made this change in the manuscript.

25: It seems like a summary-level reference might also be appropriate here (?), e.g. something from one of the recent IPCC reports (that integrates results from a large number of individual publications).
Response: Thank you for the suggestion. In addition to Aschwanden et al. (2021), we have also added the ISMIP6 papers (Goelzer et al., 2020; Seroussi et al., 2020) as reference for multi-model ensembles.

28-29: It should probably be noted here that these experiments were assuming a marine ice sheet with a significant over-deepening inland (since this configuration would necessarily be more sensitive than say, an ice sheet grounded above sea level).
Response: We have included the following in our manuscript: "While its impact at the ice sheet scale has not yet been rigorously quantified, Durand et al. (2011) showed that differences in the bed topography greater than 2 km could lead to significant variations in ice sheet model behavior for regions prone to Marine Ice Sheet Instability (MISI)."

37-38: Is the ∼2 km limit proposed here coming from the Durand et al. (2011) paper? While I more-or-less agree with this idea, I don't know that this single reference is adequate to support the precision implied by this statement. Maybe consider softening it a little bit to something less precise, e.g. "order km-scale spatial resolution"?
Response: The sentence has been modified to state "kilometer-scale spatial resolutions" instead.

63: "three regions in Greenland"; maybe add a few words of clarification here that they are glaciologically distinct / different? E.g., presumably you mean regions where velocity occurs primarily via fast sliding, a region where it occurs via a mix of sliding and deformation, etc.
Response: The sentence has been modified to state "three glaciologically distinct regions" instead.

Figure 1 caption: "The loss function is comprised of ..." or "The loss function includes data loss..."
Response: The caption has been modified to state "The loss function is comprised of data loss terms..."

89: "fully connected layers", maybe use "fully connected ('dense') layers..." ?
Response: Sentence has been modified to "fully connected ('dense') layers".

101: Should it be "the apparent mass balance residual" ?
Response: The "mass balance residual" in this case refers to the residual of the conservation of mass. The "apparent mass balance" is defined as $\dot{a} = \dot{M}_s - \dot{M}_b - \partial H/\partial t$. We have made this difference more clear in the manuscript.

103-105: I am guessing that maybe this is discussed further below (?), but it seems like you are already potentially limiting the usefulness of this approach by restricting the momentum balance to SSA. I.e., if one of the main interests here is in improving the inference in regions of slower moving ice flow, which is presumably due to less sliding and more internal deformation, then SSA doesn't seem like the right assumption to make for the model dynamics. I know that ISSM has higher-order approximations available (e.g., L1L2 or "Blatter-Pattyn"). Has that also been explored (acknowledging the obvious additional computational burden) and compared against the approach using SSA?
Response: We decided to start with a simpler approximation of the momentum balance (SSA) to minimize the computational cost and see if this method of using two conservation laws could produce sensible results. Further, since we wanted our method to be comparable to the method in BedMachine, we needed to use the depth-averaged conservation of mass and so needed to use a 2D approximation of the momentum balance. That being said, for future work in applying this approach to the slower-moving interior, we plan to use higher-order approximations of the momentum balance. We have included this in our discussion section.

116: "...from THE regional climate model RACMO..."
Response: Thank you for the suggestion, we have made this change in the manuscript.

Section 2.3: It sounds like the basal mass balance term in equation 2 is assumed to be 0? If so, it would be good to note that explicitly here in the discussion of the apparent mass balance term.
Response: We have made this change and included the following in this section: "...It should be noted that we assume the basal melting rate is small and therefore set $\dot{M}_b = 0$ m yr^{-1}..."

152: "to prevent from taking" (omit "from"?); "or diving by zero" ("dividing by zero")

Response: Thank you for catching these typos! We have made these changes in the manuscript.

176-177: Would it be worth commenting on the choice of median vs. mean? Is the median chosen because of the small number (5) of samples, such that the mean could be easily biased?
Response: Yes, we chose to use the median because we are using an ensemble of five models, and the mean could be easily biased. We have included the following in our manuscript: "It should be noted that we choose to retrieve the median prediction rather than the mean prediction since the mean of an ensemble of five models could be biased."

180: By "challenging to implement", do you mean where the traditional / previous mass conservation approach does not perform well? Implementation sounds more like the approach is challenging, but I imagine the approach is just as easy to implement in these regions, it's more the prior / baseline result that you are not happy with.
Response: By "challenging to implement", we do mean that the mass conservation inverse approach is not well constrained since the ice is moving slowly or the topography is complex (e.g., alpine topography). We have made this more clear in the manuscript.

242: It's not clear here exactly what "Fig.2(1)" is referring to.
Response: Fig. 2(l) refers to the panel in Fig. 2 depicting ice thickness measurements along ice-penetrating radar flight tracks for the Deception region. To minimize confusion, we have modified the sentence to state: "However, we also notice that PINN (MB) predicts bed features that are highly correlated to the ice-penetrating radar flight tracks in Fig. 2(l)."

Figure 5: In the caption for this and figure 4 it would be helpful to remind the reader which dataset is subtracted from which and shown in panels d-f (e.g., PINN minus original BedMachine product or vice versa?).
Response: We have made this more clear in Fig. 4 and Fig. 5. We have stated that the difference is $\hat{\boldsymbol{b}}_{\text{BedMachine}} - \hat{\boldsymbol{b}}$ (m) for panels d-f.

Table 4: It might also be useful to provide some percent / fractional metrics here? E.g., for the apparent mass balance RMSE, how does that number compare to the average apparent mass balance over the same area? Such a table could be added to the SI if it's not deemed important enough for the main text.
Response: The table has been updated to include columns for the fractional metrics (i.e., the fractional RMSEs).

3.2.2. – It's left hanging a bit as to the significance of the differences in u, apparent mass balance, and sfc. elevation when using the different approaches. For example, how do these differences compare to those that arise when using the original BedMachine approach? Would it make sense to include those metrics (differences in u, apparent mass balance, and sfc elevation) somewhere here for comparison? It's a bit unclear to me what the broader implications are of these secondary metrics w.r.t. using the derived datasets for modeling. If the authors have additional thoughts on this they would be welcome in the supplementary information.
Response: For the mass-conserving approach used in BedMachine, the training data (i.e., $\boldsymbol{v}_{\text{data}}$, $\dot{a}_{\text{data}}$) are used to directly invert for the ice thickness. With the PINN approach, to infer the ice thickness, the PINN must first predict values for the $\boldsymbol{v}, \dot{a}, s, H, C$ fields and use these within the physical loss term. Therefore, the differences in these predicted data sets arise due to the PINN architecture – the PINN can minimize error in these predicted fields by comparing them to the training data, but the predictions will never be exactly the same as the training data. We have explained this and have worked to make it more clear in Section 4 (Discussion) as BedMachine imposing mass conservation more 'strongly' and the PINN conserving the conservation laws more 'weakly'.

326: If the discussion starting in 4.1 is intended to be specific to Deception, then maybe that should be noted earlier in this paragraph? Alternatively, if the discussion in 4.1 up to line 326 where Deception is mentioned is supposed to be generic, then perhaps line 326 should be something more like, "...a far more

realistic bed topography map, particularly for Deception."
Response: The discussion section 4.1 is meant to be specific to Deception, we have made this more clear.

329-330: Would "...slightly higher RMSE SUGGESTS..." be more appropriate here than "indicates"? I think the speculation in this sentence makes sense, but it seems like it is perhaps speculation as opposed to a concrete fact.
Response: Thank you for the suggestion, we have made this change in the manuscript.

336-348: W.r.t. the prediction of thinner ice – is it also possible that this could be the result of the chosen stress balance model? E.g., in order for the SSA model to match surface velocities, it would need to assume a depth-averaged velocity profile that is larger than would be assumed in a model that allowed for internal deformation (E.g., L1L2), because SSA can only accommodate velocity via a change in the sliding component (unless I'm misunderstanding the model used here). If that is indeed the case, then it seems like the optimization process might necessarily bias the ice thickness on the thin side; if the depth averaged velocity is too large, the same flux (constrained by continuity equation and the apparent mass balance terms) can only be accommodated by reducing the ice thickness.
Response: This a good point, although the region where the PINN predicts thinner ice is generally in the fast-flowing region where we might expect there to be more sliding rather than internal deformation (which might be more typical of slower moving regions). We have added this to our discussion section in addition to the discussion of the limitation of the PINN 'struggling' to predict sharp transitions in the ice velocity.

345: "These reasons imply that..." is a bit awkward. "This implies that..."? "These arguments imply that..."?
Response: Thank you for the suggestion, we have made this change in the manuscript.

353: "...and exceeds their INDIVIDUAL limitations..."?
Response: Thank you for the suggestion, we have made this change in the manuscript.

398: "We observe that the PINN better captures..." → "We observe that the PINN captures observable features better with..."
Response: Thank you for the suggestion, we have made this change in the manuscript.

403: "...state variable predictions". (remove plural on "variable")
Response: Thank you for the suggestion, we have made this change in the manuscript.

415: "...mass-conserving approach, AS CONFIRMED BY THE DISCOVERY OF new bed features beneath Narssap and Deception."
Response: Thank you for the suggestion, we have made this change in the manuscript.

417: "...we recommend USING this approach..."
Response: Thank you for the suggestion, we have made this change in the manuscript.

A last thought / general comment: The implied "geomorphology" of the three focus areas studied here look very different from one another. E.g., The Upernavik and Narsaap beds look very smooth when compared to Deception. In the areas where there are no troughs, they almost look like high-resolution DEMs from past, heavily glaciated regions of Canada. Is there any published work on previous Greenland glaciations that might provide some more insight into this? I'm not suggesting it should be part of this paper, but it could be interesting to look into whether or not the "smoothness" that your methods are implying about the bed in different regions is in line with current glacial geological / geomorphological understanding. It would seemingly be a further testament to the power of the methods used here if you were resolving that level of information about the bed through hundreds / thousands of meters of ice.
Response: Thank you for your comment. We agree that the geomorphology of the regions studied in this paper look quite different from each other, and it would indeed be interesting to understand whether the PINN predicted beds are in line with current glacial geological / geomorphological understanding as part

of a future project. For the purposes of this paper, when we refer to “smooth” PINN predictions, we are comparing predictions directly with the data along ice penetrating radar flight tracks which we observe is far more “rough”. Like BedMachine, the PINN also does capture lower-frequency approximations of the bed, however misses out on the higher-frequency details.

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.
Response: We appreciate the reviewer's thoughtful comments. Thank you for your positive feedback and for highlighting the relevance of our methodological study.

I have a few comments, summarized below:

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 our validation study. Regarding the point about the uncertainty, we have tried to be careful about comparing the predictions and the training since we acknowledge that a small RMSE between predicted and training data does not necessarily imply a good prediction between tracks. However, it is important to note that the PINN is not exposed to all of the available ice thickness measurements during training. Rather, the training process involves randomly selecting a small subset of the ice thickness measurements, so comparing all the available $\hat{\boldsymbol{b}}_{\text{data}}$ to the predicted $\hat{\boldsymbol{b}}$ along the flight tracks is reasonable from this perspective. To further address your point, we have expanded our validation section in the appendix and included two more experiments to test the effectiveness of PINN predictions by changing the track spacing as well as validating against independent ice thickness data.

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 the Shallow-Shelf Approximation (SSA) to estimate the future sea level contribution of the Greenland Ice Sheet. That being said, we do acknowledge that this assumption is a limitation of our approach and have expanded on this in the discussion section. For this paper, 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, higher-

order modeling will be necessary to apply this approach to the interior, slower-moving region of the ice sheet.

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: 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* the mass conserving domain of BedMachine, $\Omega \backslash \Omega_{\mathrm{mc}}$, as stated in the manuscript. In $\Omega \backslash \Omega_{\mathrm{mc}}$, BedMachine uses other interpolation methods like kriging or streamline diffusion (not mass conservation), whereas PINN (MB), PINN (SB) are attempting to constrain predictions to the conservation of mass and conservation of momentum respectively. In these two cases, we think the PINN is "overfitting" to the data rather than learning from the physical loss terms. This improves with PINN (MB+SB) as we see adding more physical loss terms (i.e., adding more regularization terms) provides the PINN with more information, leading to a more realistic prediction.

Regarding the difference between the input data, BedMachine is exposed to all the available ice thickness data, and it numerically solves for the ice thickness in between flight tracks using velocity and apparent mass balance data. The PINN is exposed to a small, randomly-sampled, subset of all the available ice thickness data (i.e., we choose 4000 ice thickness data points from the available ice thickness data for training); these data are then used to constrain predictions within the region of interest. We have made this more explicit in our discussion section and have cited the BedMachine Greenland paper with all the technical details (Morlighem et al., 2017).

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 have included an additional figure in the Appendix of the mass balance and momentum balance residuals. Note that the PINN will not satisfy the physics perfectly since it is solving a minimization problem. It should also be noted that all of our predicted maps (including the predicted bed topography, velocity, apparent mass balance, surface elevation, basal friction coefficient, and PDE residual maps) have an adaptive mesh resolution of 400-500 m. Therefore, all of the predictions are evaluated on ~25000-42000 (x, y) points depending on the size of the region of interest.

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 have included a figure that depicts the standard error 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 the traditional adjoint method for a flow-line SSA problem can be found in Cheng et al. (2021), however the numerical implementation of a 2D problem would require automatic differentiation. We have elaborated on this in the methods and discussion sections of the manuscript.

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: Thank you for the suggestion, we have made this change in the manuscript.

Line 85: "satisfy the PDE residuals" → "satisfy the PDEs"
Response: Thank you for the suggestion, we have made 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 have re-made 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 resampling the collocation points during the training process and is a feature built-in to PINNICLE (Cheng et al., 2025), it can add to the computational cost because resampling often leads to slower convergence. We have chosen a higher number of collocation points for this reason, so that the physical laws are generally conserved throughout the region of interest. Moreover, we have trained an ensemble of five 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 PINNs satisfying the physics for fixed collocation points. Indeed, Figure 1 (above) shows the coverage of data points, the number of which are lower than the number of collocation points for each PINN. We then retrieve the median prediction of the five PINNs. We have made this clear in the methods section.
Further, we added another section in our Appendix, with the results of an experiment of a PINN trained with resampling of collocation points every 100 iterations. Since we are resampling the collocation points, we reduce the number of collocation points to 4000 (instead of 9000). We find that this result is similar to the results of our PINN (MB+SB) ensemble.

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 the feedback. We have made 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/yr 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. 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. The regions have different velocities, but 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 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. In the manuscript, we acknowledge that there could be some variability or flexibility in choosing these weighting terms since the velocities are different, and this might be something we test in the future for slower-moving ice in the interior of the Greenland Ice Sheet. We have made this more clear in our 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. We have provided more explanation for these choices in the methods section.

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: Thank you 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 v_x and 4000 data points for 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 have made 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 $\hat{\boldsymbol{b}}_{\text{BedMachine}} - \hat{\boldsymbol{b}}$ (m) 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 the feedback. We have included an additional figure, Fig. 6, that depicts the standard error of the PINN ensemble for each region, as well as the corresponding statistics in Table 4.

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: While it is possible that we are excluding some of the information by taking the median across all PINN predictions, based on the results in Fig. 6a, the standard error in PINN (MB+SB) ensemble for Upernavik depicts that the predictions for the area beneath the southern fork of Upernavik Isstrøm North generally converge, implying that majority of the PINNs in the ensemble predict a 'disconnected' trough. That being said, the 'disconnected' part of the trough is shaded dark green which is very close to a bed elevation of 0 m (Fig. 5a), indicating that while the PINN does not predict the full depth of the trough (as shown in BedMachine), it is quite close. We have added this to our discussion.

Table 4: Why is thickness RMSE not included?

Response: Since the bed topography, b is calculated by subtracting the thicknesss values from surface elevation values, $b = s - H$, the thickness RMSEs are effectively the same as the bed topography RMSEs. These are included in Tables 2 and 3. Table 5 (previously Table 4) is only meant for the other state variables in the PDEs for which we have training data (i.e., velocity, apparent mass balance, and surface elevation).

References

Aschwanden, A., Bartholomaus, T. C., Brinkerhoff, D. J., and Truffer, M.: Brief communication: A roadmap towards credible projections of ice sheet contribution to sea level, Cryosphere, 15, 5705–5715, https://doi.org/10.5194/tc-15-5705-2021, 2021.

Cheng, G., Kirchner, N., and Lötstedt, P.: Sensitivity of ice sheet surface velocity and elevation to variations in basal friction and topography in the full Stokes and shallow-shelf approximation frameworks using adjoint equations, The Cryosphere, 15, 715–742, https://doi.org/10.5194/tc-15-715-2021, 2021.

Cheng, G., Morlighem, M., and Francis, S.: Forward and Inverse Modeling of Ice Sheet Flow Using Physics-Informed Neural Networks: Application to Helheim Glacier, Greenland, Journal of Geophysical Research: Machine Learning and Computation, 1, https://doi.org/https://doi.org/10.1029/2024JH000169, 2024.

Cheng, G., Krishna, M., and Morlighem, M.: A Python library for solving ice sheet modeling problems using Physics Informed Neural Networks, PINNICLE v1.0, EGUsphere, 2025, 1–26, https://doi.org/10.5194/egusphere-2025-1188, 2025.

Durand, G., Gagliardini, O., Favier, L., Zwinger, T., and le Meur, E.: Impact of bedrock description on modeling ice sheet dynamics, Geophysical Research Letters, 38, https://doi.org/10.1029/2011GL048892, 2011.

Goelzer, H., Nowicki, S., Payne, A., Larour, E., Seroussi, H., Lipscomb, W. H., Gregory, J., Abe-Ouchi, A., Shepherd, A., Simon, E., Agosta, C., Alexander, P., Aschwanden, A., Barthel, A., Calov, R., Chambers, C., Choi, Y., Cuzzone, J., Dumas, C., Edwards, T., Felikson, D., Fettweis, X., Golledge, N. R., Greve, R., Humbert, A., Huybrechts, P., Le Clec'h, S., Lee, V., Leguy, G., Little, C., Lowry, D. P., Morlighem, M., Nias, I., Quiquet, A., Rueckamp, M., Schlegel, N.-J., Slater, D. A., Smith, R. S., Straneo, F., Tarasov, L., van de Wal, R., and van den Broeke, M.: The future sea-level contribution of the Greenland ice sheet: a multi-model ensemble study of ISMIP6, Cryosphere, 14, 3071–3096, https://doi.org/10.5194/tc-14-3071-2020, 2020.

Morlighem, M.: MEaSUREs BedMachine Antarctica, Version 4, https://doi.org/10.5067/POJQI54A45HX, 2025.

Morlighem, M. and et al: IceBridge BedMachine Greenland, Version 6, https://doi.org/10.5067/6B6B225B8V2D, 2025.

Morlighem, M., Rignot, E., Seroussi, H., Larour, E., Ben Dhia, H., and Aubry, D.: A mass conservation approach for mapping glacier ice thickness, Geophysical Research Letters, 38, https://doi.org/10.1029/2011GL048659, 2011.

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., Catania, G., Chauche, N., Dowdeswell, J. A., Dorschel, B., Fenty, I., Hogan, K., Howat, I., Hubbard, A., Jakobsson, M., Jordan, T. M., Kjeldsen, K. K., Millan, R., Mayer, L., Mouginot, J., Noel, B. P. Y., O'Cofaigh, C., Palmer, S., Rysgaard, S., Seroussi, H., Siegert, M. J., Slabon, P., Straneo, F., van den Broeke, M. R., Weinrebe, W., Wood, M., and Zinglersen, K. B.: BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation, Geophysical Research Letters, 44, 11 051–11 061, https://doi.org/10.1002/2017GL074954, 2017.

Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., Eisen, O., Ferraccioli, F., Forsberg, R., Fretwell, P., Goel, V., Greenbaum, J. S., Gudmundsson, H., Guo, J., Helm, V., Hofstede, C., Howat, I., Humbert, A., Jokat, W., Karlsson, N. B., Lee, W. S., Matsuoka, K., Millan, R., Mouginot, J., Paden, J., Pattyn, F., Roberts, J., Rosier, S., Ruppel, A., Seroussi, H., Smith, E. C., Steinhage, D., Sun, B., van den Broeke, M. R., van Ommen, T. D., van Wessem, M., and Young, D. A.: Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the Antarctic ice sheet, Nature Geoscience, 13, 132–137, https://doi.org/10.1038/s41561-019-0510-8, 2020.

Otosaka, I. N., Shepherd, A., Ivins, E. R., Schlegel, N.-J., Amory, C., van den Broeke, M. R., Horwath, M., Joughin, I., King, M. D., Krinner, G., Nowicki, S., Payne, A. J., Rignot, E., Scambos, T., Simon, K. M., Smith, B. E., Sørensen, L. S., Velicogna, I., Whitehouse, P. L., A, G., Agosta, C., Ahlstrøm, A. P., Blazquez, A., Colgan, W., Engdahl, M. E., Fettweis, X., Forsberg, R., Gallée, H., Gardner, A., Gilbert, L., Gourmelen, N., Groh, A., Gunter, B. C., Harig, C., Helm, V., Khan, S. A., Kittel, C., Konrad, H., Langen, P. L., Lecavalier, B. S., Liang, C.-C., Loomis, B. D., McMillan, M., Melini, D., Mernild, S. H., Mottram, R., Mouginot, J., Nilsson, J., Noël, B., Pattle, M. E., Peltier, W. R., Pie, N., Roca, M., Sasgen, I., Save, H. V., Seo, K.-W., Scheuchl, B., Schrama, E. J. O., Schröder, L., Simonsen, S. B., Slater, T., Spada, G., Sutterley, T. C., Vishwakarma, B. D., van Wessem, J. M., Wiese, D., van der Wal, W., and Wouters, B.: Mass balance of the Greenland and Antarctic ice sheets from 1992 to 2020, Earth System Science Data, 15, 1597–1616, https://doi.org/10.5194/essd-15-1597-2023, 2023.

Palmer, S., Kirkwood, C., Luo, C., and Morlighem, M.: A quantile regression forest estimate of Greenland's subglacial topography, Journal of Glaciology, 71, e115, https://doi.org/10.1017/jog.2025.10071, 2025.

Pritchard, H. D., Fretwell, P. T., Fremand, A. C., Bodart, J. A., Kirkham, J. D., Aitken, A., Bamber, J., Bell, R., Bianchi, C., Bingham, R. G., Blankenship, D. D., Casassa, G., Christianson, K., Conway, H., Corr, H. F. J., Cui, X., Damaske, D., Damm, V., Dorschel, B., Drews, R., Eagles, G., Eisen, O., Eisermann, H., Ferraccioli, F., Field, E., Forsberg, R., Franke, S., Goel, V., Gogineni, S. P., Greenbaum, J., Hills, B., Hindmarsh, R. C. A., Hoffman, A. O., Holschuh, N., Holt, J. W., Humbert, A., Jacobel, R. W., Jansen, D., Jenkins, A., Jokat, W., Jong, L., Jordan, T. A., King, E. C., Kohler, J., Krabill, W., Maton, J., Gillespie, M. K., Langley, K., Lee, J., Leitchenkov, G., Leuschen, C., Luyendyk, B., MacGregor, J. A., MacKie, E., Moholdt, G., Matsuoka, K., Morlighem, M., Mouginot, J., Nitsche, F. O., Nost, O. A., Paden, J., Pattyn, F., Popov, S., Rignot, E., Rippin, D. M., Rivera, A., Roberts, J. L., Ross, N., Ruppel, A., Schroeder, D. M., Siegert, M. J., Smith, A. M., Steinhage, D., Studinger, M., Sun, B., Tabacco, I., Tinto, K. J., Urbini, S., Vaughan, D. G., Wilson, D. S., Young, D. A., and Zirizzotti, A.: Bedmap3 updated ice bed, surface and thickness gridded datasets for Antarctica, Scientific Data, 12, 414, https://doi.org/10.1038/s41597-025-04672-y, 2025.

Sandberg Sørensen, L., Simonsen, S. B., Forsberg, R., Stenseng, L., Skourup, H., Savstrup Kristensen, S., and Colgan, W.: Circum-Greenland, ice-thickness measurements collected during PROMICE airborne surveys in 2007, 2011 and 2015, GEUS Bulletin, 41, 79–82, https://doi.org/10.34194/geusb.v41.4348, 2018.

Seroussi, H., Nowicki, S., Payne, A. J., Goelzer, H., Lipscomb, W. H., Abe-Ouchi, A., Agosta, C., Albrecht, T., Asay-Davis, X., Barthel, A., Calov, R., Cullather, R., Dumas, C., Galton-Fenzi, B. K., Gladstone, R., Golledge, N. R., Gregory, J. M., Greve, R., Hattermann, T., Hoffman, M. J., Humbert, A., Huybrechts, P., Jourdain, N. C., Kleiner, T., Larour, E., Leguy, G. R., Lowry, D. P., Little, C. M., Morlighem, M., Pattyn, F., Pelle, T., Price, S. F., Quiquet, A., Reese, R., Schlegel, N.-J., Shepherd, A., Simon, E., Smith, R. S., Straneo, F., Sun, S., Trusel, L. D., Van Breedam, J., van de Wal, R. S. W., Winkelmann, R., Zhao, C., Zhang, T., and Zwinger, T.: ISMIP6 Antarctica: a multi-model ensemble of the Antarctic ice sheet evolution over the 21st century, Cryosphere, 14, 3033–3070, https://doi.org/10.5194/tc-14-3033-2020, 2020.