We thank both anonymous referees and the editor for their time and care in reading our manuscript and for their encouraging comments and constructive feedback. The manuscript has been updated to reflect the referees' comments.

We note additional minor changes to the manuscript:

- The references to the preprint by Schmitt et al. (2025) have been updated to reference the published paper by Schmitt et al. (2026).
- The references to numbered equations throughout the text have been standardized in line with the journal's submission guidelines.
- Minor typographical errors have been corrected.

Our responses to the referees' specific comments are provided below.

Author response to RC1

Referee comments in blue
Author comments in black

General comments

> With this manuscript, the authors present a new estimate of the thickness of a glacier in High Mountain Asia (the Bhote Kosi glacier) using a high-order, deep-learning-based model (the Instructed Glacier Model, or IGM) that improves upon existing, less reliable low-order models. The primary broader goal of this effort is to establish an optimized case study that can then be applied to thickness estimates of other mountain glacier systems. The authors develop an inversion method to derive glacier thickness from ice surface velocity alone, then validate this thickness estimate with previously published thickness observations derived from airborne radar. Finally, the authors investigate the role that the spatial distribution of thickness observations plays in the calibration. The authors ultimately find that the velocity-only inversion with IGM is generally successful in reproducing the ice thickness field for Bhote Kosi glacier with errors comparable to the global thickness inversion estimates of Millan et al. (2022). Notably, the higher-order physics of IGM improve upon the Millan study for one branch of the Bhote Kosi glacier, potentially by including non-negligible stress components that are omitted in the Shallow Ice Approximation.
>
> This is a well-motivated, well-written manuscript that makes an interesting contribution to the growing literature on deep-learning techniques applied to mountain glacier settings. The discussion of relevant parameters, results, implications for other glaciers, and recommendations for future observation locations is generally thorough and thoughtful. I found the IGM description and methodology to be relatively accessible and well explained. My suggestions for improvement are minor.
>
> I recommend the manuscript be accepted for publication pending the authors' attention to the comments below at their discretion.

Thank you for your thoughtful response and your constructive comments which will help to improve the manuscript.

Specific comments

> 1. Although the description of the development of ice thickness estimation methods (lines 62–102) in the Introduction is informative, it reads as somewhat encyclopedic and could be made more concise.

Taking this comment into consideration, along with the comment from reviewer #2 that the Introduction is educational for readers with less knowledge of mountain glacier research, we will make these paragraphs more concise in such a way that is intended to convey the same information.

> 2. Although the choice of α_h among $\alpha_h = 2, 3, 6, 10$ does not substantively affect the MAE value, the authors note a strong volume dependence on α_h. The authors should therefore consider presenting results in the Appendix of one or more of the key experiments specifically with $\alpha_h = 2, 3, 10$, not just $\alpha_h = 6$, for the sake of completeness. The authors also note that valid arguments could be made for choosing any one of $\alpha_h = 2, 3, 6, 10$, so it would be beneficial to include a brief discussion of the arguments favoring each one of these four particular values.

Qualitative differences between the results for $\alpha_h = 2, 3, 6, 10$ are minimal, and quantitative differences integrated across the glacier are already shown in Figure 2. We have chosen not to include the resulting thickness maps for each in the manuscript, as they look similar and we feel it would add little value. However, we show them in Figure AR1-1 in response to this comment.

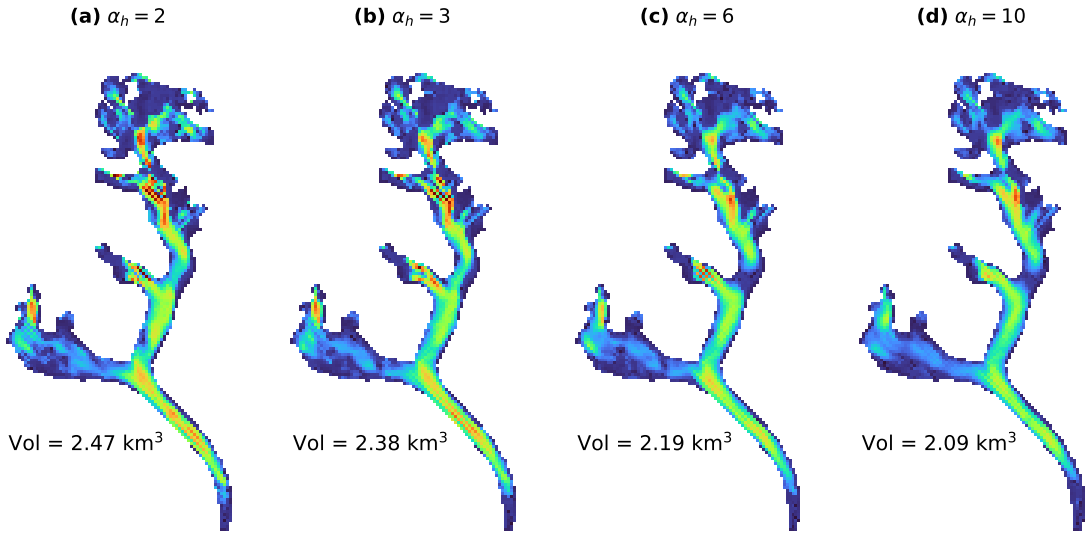


Figure AR1-1: Thickness maps of Bhote Kosi glacier for four different choices of the thickness regularization parameter α_h: (a) $\alpha_h = 2$, (b) $\alpha_h = 3$, (c) $\alpha_h = 6$, (d) $\alpha_h = 10$. The colour scale for all panels is identical to that of Fig. 6(a,b) in the manuscript.

The "valid arguments" paragraph will be reworded for clarity. It is not that there are entirely separate arguments for choosing each of the α_h, but that the "best" value can be chosen taking multiple factors into account (regularization cost, velocity cost, MAE, MBE, and total volume). As

the L-curve is a heuristic method, any of the values $\alpha_h = 2$, 3, 6, 10 are in some sense valid, and the choice of any one of them over the others is subjective.

> 3. The biases in Figure 3 (namely: thickest-parts, flattest-parts, and low-elevation) would be made clearer if presented quantitatively. That is, it would be useful to give the readers a quantitative sense of the observed thicknesses, surface slopes, and elevations in question within the same figure. This information could potentially be visualized in a map overlay, or in a separate row of subplots.

The thickest-parts and flattest-parts constraining sets were chosen by sorting the 23 profiles by their maximum thickness and mean slope respectively, and the top 5 or 6 profiles were chosen as the constraining set. The low-elevation constraining set was chosen by selecting profiles with at least one point in the lowest 10% of the glacier's elevation range. For each constraining set, profiles intersecting with profiles already in the constraining set were also included. Figure 3 will be updated to show the ranges of observed thicknesses, surface slopes, and elevations considered in each constraining set.

While updating Figure 3, it was noted that the maximum observed thickness on Bhote Kosi was incorrectly stated later in the text as 287 m. The text will be updated to reflect the correct value, which is 292 m.

> 4. The title of the manuscript is somewhat generic and slightly undersells the contribution of this work. I suggest the authors consider changing the title to be more informative, perhaps emphasizing the study's result of deriving glacier thickness from velocity using a deep-learning-based modeling framework.

Thank you for this suggestion. The title of the manuscript will be amended to reflect the role of the neural network emulator in our methodology.

Technical corrections

> [Line 70] The comma after "ice thickness" should be removed.

> [Line 104] The comma after "glacier model" should be removed.

> [Line 126] Only the acronym "IGM" is needed here, as it was already defined in the Introduction.

> [Line 132] Only the acronym "CNN" is needed here, as it was already defined in the previous paragraph.

> [Line 229] The comma after "dataset" should be removed.

> [Line 237] The word "performing" should be "perform".

> [Figure 3 caption] To maintain consistency with other figure captions, "partitions" after (a–d) and "mean" after (e–h) should be capitalized.

Thank you for these technical corrections; the text will be updated as requested.

Author response to RC2

> This manuscript presents an in-depth case study of glacier ice thickness inversion for the Bhote Kosi glacier in the Himalaya, which is performed using the deep-learning-based Instructed Glacier Model (IGM). The authors invert the observed surface velocities for the ice thickness, and the L-curve technique is utilized as to selecting the regularization parameter. The results are validated against the recent airborne radar thickness observations of Pritchard et al. (2026). Furthermore, cross-validation experiments with thickness-constrained inversions have been performed, and it is found that the thickness observations correct the modelled field locally, especially in the low-velocity areas, however they have a weak interpolative power at the unconstrained locations. The study is concluded with practical recommendations as to the choice of the inversion parameters and the placement of the future thickness measurements.

> It is noteworthy that the manuscript is written very well.

Thank you for your thoughtful response and your constructive comments which will help to improve the manuscript.

> The introduction has a very good flow and brings even the readers who are not so close to the mountain glacier research up to speed with the topic.

Thank you for this comment; this was our intention. In response to a comment from reviewer #1, the part of the Introduction concerning the existing literature on glacier thickness estimation will be made slightly more concise in such a way that is intended to convey the same information.

> Line 144: it is stated that the approach differs from PINNs in two ways, however these differences are not explained clearly. A bit more clarification would be of help here.

This paragraph will be updated to expand on what is typically meant by a PINN, and why IGM is different from the original PINN method. Furthermore, throughout the manuscript we will instead refer to the neural network "loss" rather than the "cost", as this terminology is more common when referring to neural networks, avoids confusion with the optimization cost function discussed elsewhere in the text, and hopefully aids in understanding of how IGM relates to PINNs.

> It is understood that the model is based on the work of Jouvet and Cordonnier (2023), however the reader should be able to fully understand the approach and the method without having read that publication. A more standalone explanation of the relevant details from that work would be appreciated.

The formulation of the energy functional of the Blatter-Pattyn model (which acts as the neural network loss) is omitted from the manuscript as it is not directly relevant to this work. If it were included, we believe readers may confuse it with the cost function (Eq. 2). We will add a sentence clarifying where this can be found elsewhere in the literature. We will also include a paragraph at the end of Sec. 2.1 to clarify that the optimization method does not require the velocity solve to be emulated by a neural network.

> It is not clear which of the two following aspects makes the model a PINN: (1) the fact that the weights λ of the CNN are trained to minimize the energy functional associated with the Blatter-Pattyn model, or (2) the fact that the aim is to obtain an optimal fit of the control variables across the whole glacier, in such a way that the misfit between the surface velocity output from the emulator (Eq. 1) and the observed surface velocity is small (Jouvet, 2023). Some clarification on this point would be of use.

(1) is exactly what makes IGM's ice flow emulator a (variational) PINN, and the text will be updated to make this clearer.

(2) is the way in which our thickness inversion method is set up as an optimization problem. This does not require the emulator to be a PINN, but encoding the ice flow physics within a neural network makes the optimization framework computationally efficient, simple to set up, and allows us to use automatic differentiation to find gradients of the optimization cost with respect to the thickness field. As noted in response to your previous comment, we will include a paragraph at the end of Sec. 2.1 to clarify this.

> Line 191: it is not clear why the pre-trained model undergoes only one iteration of retraining and not more. A brief justification would be useful.

While it would be possible to implement a different retraining schedule (e.g. more iterations), a proper investigation of the effect of this was beyond the scope of this paper. We used one retraining iteration per inversion iteration as it was the model default, and a brief justification will be included.

> The authors have shown that the lack of thickness observations need not be a barrier to producing a reasonable thickness estimate, which is a very important result and is nicely applicable to the other glaciers around the world. However, it is also acknowledged that the method is unable to produce similar results for the glaciers with no measured thickness. Demonstrating this limitation with an example would provide a valuable proof of the limitations of the approach and would add to the completeness of the paper.

To clarify, in the absence of measured thicknesses, it is not possible to perform a validation test. One can produce an L-curve in the style of Figure 2(a) and select a regularization parameter based on that, but there is no way to calculate the error against the true thickness, i.e. there is no way to make the equivalent of Figure 2(b). We will clarify the text here, but we do not feel that demonstrating this limitation with an example would be any more effective than describing it in the text.

The method was applied with the same regularization parameter identified here ($\alpha_h = 6$) to other glaciers for which thickness observations were available, and the results were of mixed quality. This indicates that further work may be required to find the ideal regularization parameter for each individual glacier. Results were not shown, as we feel they would add little value to this publication.

> The importance which is given to advising on the observation locations for the low-velocity glaciers is very significant and addresses a gap in the available observed data.

> The inversion is initialized with the thickness estimate of Millan et al. (2022) (Sect. 2.2.1), while the observed velocities also come from the same study. In Sect. 4.3, the shared velocity data is acknowledged, however the shared initialization is not: due to the fact that the cost function is non-convex, the final thickness field might retain some dependence on the initial guess. A sensitivity test with a different initialization (e.g. the consensus estimate, or a uniform thickness) would confirm that the result is independent of the starting point and would strengthen the comparison in Fig. 6.

Thank you for this comment; this is indeed something that was not discussed. We will include a paragraph describing the sensitivity test that was performed, and include the parameter file for this test in the data repository.

> There is a temporal mismatch between the input datasets: the DEM is from 2010–2015, the velocities are from 2017–2018, and the thickness observations are from late 2019. It is mentioned in the conclusions that the contemporaneity between the input datasets is important, however this is not assessed for Bhote Kosi itself. The surface lowering between the DEM epoch and the radar survey could bias the comparison with a positive expected sign, which is likely still very small. A brief statement of the expected sign and magnitude of this effect would be valuable.

Thank you for pointing this out. We did raise the issue of contemporaneity of data, but this was not mentioned prior to the conclusion and no specifics were given. Assuming a constant $\mathrm{d}h/\mathrm{d}t$ of around 2 m yr^{-1} (Hugonnet et al., 2021), we estimate the total surface lowering between the DEM date and 2019 to be no more than 20 m. As we do not correct for this lowering, this introduces a small positive bias in comparison with the Pritchard et al. (2026) data, but notably with a magnitude much smaller than the misfit we observe (see e.g. Fig. 8(b)). As such, we feel that although we did not correct for surface lowering, it does not bias our results. We will give clearer detail about these issues in our revision, and introduce them in the discussion, rather than the conclusion.