Author response to RC2

Referee comments in blue
Author comments in black

> 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.