Response letter

Referee comments in black.
Author comments in green.

Response to Reviewer #1

We thank the reviewer for another evaluation of our manuscript and are glad to hear that all comments were satisfactorily addressed.

Response to Reviewer #2

The authors present a revised manuscript addressing the comments from the initial review round. Although I appreciate the authors effort to answer the major comments, I regret to say that some of their answers to several of my major concerns are not convincing - at least to me. I still remain very positive about this manuscript and therefore recommend that the editor should continue to considered it for publication in The Cryosphere after my concerns have been alleviated.

We are grateful to the reviewer for taking the time and effort to go through the manuscript again. We are glad the reviewer is still positive about the manuscript and regret that we were not able to take away all the concerns with our previous responses. With the new revisions and replies below we hope to have satisfactorily addressed the remaining concerns!

MAJOR COMMENTS

METHODOLOGICAL MIX
Thank you for including the comparison between PISM and IGM including RMSE values and thickness maps in the rebuttal (not in the manuscript). You argue that PISM performs better for regional application. The fact why you prefer IGM over PISM for small-scale glaciers remains vague (higher-order) and not convincing to me. I still do not see the ultimate argument to included IGM - certainly after you invoke the limitations of the current fast developments. I still wonder why you are not more consistent and apply PISM all over the domain (with prescribed perfect-plasticity values for surging glaciers). In this way, you would get a more coherent map product and a simpler method (also in terms of calibration).

Thank you for this comment. After the first review round we mistakenly assumed the reviewer was only wondering why we did not use IGM everywhere (possibly because of the focus on that in the other review), and hence did not discuss the option whether to use PISM everywhere. In our previous response we addressed why PISM is preferred over IGM for the large (tidewater) glaciers, which is primarily because IGM has not yet been tested extensively on such glaciers. And we found worse performance of IGM for these glaciers in a first test in northwestern Svalbard.. For small land-terminating glaciers we rather find the opposite. IGM has in various applications on (small) mountain glaciers shown strong performance (e.g. Cook et al. 2023; Jouvet and Cordonnier, 2023; Jouvet et al. 2022) and its low computational cost and higher-order physics make it a suitable model for high-resolution modeling of small (steep) glaciers. We argue from an ice flow physics point of view that the benefit of using higher-order physics for small (often steeper) mountain glaciers is larger than for larger thick (relatively flat) tide-water glaciers and ice caps. This is because the shallowness assumptions (SIA and SSA) become less accurate for the typically larger depth-to-width ratios of mountain glaciers. Besides this, our main argument to use IGM for small glaciers is that it is simply performing better than PISM for these glaciers. We have modeled the small glaciers also with PISM (at 500-m resolution) and a comparison of the related statistics reveals that using PISM instead of IGM leads to an increase of the MAE from 38.0 to 42.7 m and RMSE from 50.1 to 54.1 m. Furthermore, the R-correlation drops from 0.77 (IGM) to 0.71 (PISM). To allow for a direct comparison we have first reprojected the 500-m PISM results to the finer 100-m grid used by IGM using nearest-neighbor interpolation.

We agree with the reviewer that the use of two instead of one ice flow model reduces the coherency of the results of glaciers in classes 1 and 2. However, our aim is to produce a best possible thickness map, which as shown requires the use of different flow models. In a similar fashion we justify the use of two different resolutions for glaciers in class 1 (100-m) and class 2 and 3 (500-m). This also reduces coherency between the results of the two classes but we value the greater detail in the final thickness and bed maps more than the downside of having inconsistent resolutions across the thickness and bed maps.

We reformulated and added the following in Sect. 4.2 (3rd paragraph):
"*It is noteworthy that in case PISM was used for the glaciers currently modeled with IGM (class 1), the MAE would increase to 42.7 m (IGM: 38.0 m), RMSE to 54.1 m (IGM: 50.1 m) and R would drop to 0.71 (IGM: 0.77). For this comparison, PISM results on the 500-m resolution grid were reprojected to the 100-m resolution IGM grid using nearest neighbor interpolation. The above confirms that the use of IGM for small glaciers leads to better agreement with thickness measurements. One reason may be the higher-order physics behind IGM, which helps to resolve small-scale ice flow and bed features better than with a model like PISM which is based on shallowness assumptions (i.e. small depth-to-width ratios are less likely to apply to glaciers in class 1).*"

In Sect. 4.3 (first paragraph) the following was added/reformulated:
"*By applying dedicated inverse methods and model physics for different glacier types, using state-of-the-art remote sensing and model input datasets, and calibrating against thickness observations, we limit uncertainties in the final thickness and bed maps. Arguably, using different ice flow models, spatial resolution, and individual parameter calibration per glacier class, causes some consistency between the methods to be lost. However, advantageously we achieve a lower misfit with thickness observations by treating glacier types separately. More specifically, the superior performance of IGM for glaciers in class 1, as well as the improved results with PISM for glaciers in class 2, were the main reasons to use two different ice flow models for these classes.*"

and

"*In summary, our modelling choices led to more detailed bed and thickness maps that are in closer agreement with observations, yet at the expense of some coherency.*"

PROGRESS

In response to this comment, you raise the argument that your approach 'can be used as a numerically stable spin up state for prognostic modeling'. I agree that this initialisation is big asset. Yet the built-in IGM inversion for thickness and ice-flow parameters claims the same property. Please discuss. Do not misunderstand me, I can accept this argument as a clear benefit. Yet, I imagine myself to start modelling on Svalbard with your thickness map. I would then need to use IGM and PISM to do so consistently. It might be me, but it seems impractical to do so with two models that need to communicate/interact. Again, the solution is to reduce the method mix to PISM-only (with perfect plasticity).

Thanks for this comment. Admittedly, it would be more work to perform spin-up and subsequent prognostic modeling with two models rather than with one model (more or less twice the amount of work). In the way we have currently chosen to categorize the glaciers, with glaciers in class 1 (small land-terminating glaciers) not having shared boundaries with glaciers in class 2 (large glaciers and ice cap systems) we however disagree that communication or interaction is necessary between the models. One problem we do envision though is for glaciers in class 3 (surging glaciers). Since neither PISM or IGM was used for those glaciers, no spin-up has been done for those glaciers and starting a forward simulation in the present day with an arbitrary ice flow model for those glaciers would likely lead to a 'shock', i.e. sudden geometric changes, at the start of a prognostic run. We have currently no suitable solution for this other than trying to use either PISM or IGM instead for these glaciers as well when preparing for a future simulation.

We added the following to Sect. 4.2 (final paragraph):
"*... has the advantage that it can be used as a numerically stable spin up state for prognostic modeling. This currently however only applies to glaciers in classes 1 and 2, for which iterative inverse methods were used. In case also glaciers in class 3 are to be included in a prognostic run, we would suggest to instead use PISM also for these glaciers to allow for spin up and transient forward modelling (as for glaciers in class 2). This inevitably does introduce larger uncertainty in the basal topography and initial ice thickness.*"

And yes, it is correct that the built-in IGM inversion also can be used to generate a stable spin-up state for prognostic modelling. Similarly, several other inversion methods can be set up to do so as well (e.g. Farinotti et al. 2009) if one aligns the prognostic modelling methodology with the inversion workflow (e.g. same assumptions on ice flow physics, same grid). However, among the so-far existing thickness products for Svalbard, none have been generated in that fashion until now. For example, the Fürst et al. (2018) thicknesses in fast flowing areas were derived using mass-conservation and observed velocities (without any ice flow model), meaning that these thicknesses cannot be used for prognostic simulations without spin-up. This is why we name this benefit in sect. 4.2. With that said, we do not think that it would add much to include more on how other inversion methods could be set up to achieve the same goal.

PERFECT PLASTICITY

Your answers to my minor comments on L91 and L107-108 are not yet satisfying. You agreed that temporal consistency motivated your choice - specifically with regard to the 2010

DEM and 2017-2018 surface velocity observations. Following this logic, you should rather use the Copernicus DEM. Moreover, the perfect plasticity does not require surface velocities. So your argument should invoke the glacier outlines and the DEM. This also corroborates your argument on limiting the surge-type identification to the velocity fields from 2017-18. As I indicated before, the analysis of Koch et al. (2023) shows an extended time coverage for surge-type glacier identification. Finally, you did not comment on the applicability of the perfect plasticity approach for surge-type glaciers. Please add.

We are grateful for these comments. Temporal consistency was one of the factors considered for input data selection. Another one was the quality of the input data. For the DEM, we chose the NPI S0 Terrengmodel which is a dedicated product for Svalbard based on aerial photos and with high spatial resolution. More details are available here: https://data.npolar.no/dataset/dce53a47-c726-4845-85c3-a65b46fe2fea. We used the 20-m resolution Svalbard-wide model here, which is derived from sub-region models at 2-5 m spatial resolution. This DEM is widely used in many studies on Svalbard, e.g. to quantify surface height change (1936-2010) in a recent study in Nature by Geyman et al. (2022). Other DEMs, such as the Copernicus DEM and the ArcticDEM have not been specifically optimized for Svalbard. This was the main reason for us to choose the NPI DEM. Besides that, the highest resolution Copernicus DEM for Europe (EEA-10) does unfortunately not include Svalbard in its domain, hence we would have to use a global version (GLO-30 or GLO-90) instead which has coarser spatial resolution (30 or 90 m). Furthermore, the Copernicus DEM is from 2011-2015, which would still be a 4-5 year mismatch with the timing of the velocity dataset. The following was added to Sect. 2:
"*For surface heights, we chose to use the S0 Terrengmodel by the Norwegian Polar Institute (NPI, 2014), which is a 20-m resolution digital elevation model (DEM), based on aerial photos between 2009-2012 and derived from subset models (5-m resolution) for regions in Svalbard.*"

Our main aim was to select glaciers that due to surging would be affected strongly by the mismatch in timing of the input datasets (velocity, dh/dt and DEM). We use the timing of the velocity dataset to select which glaciers are problematic to model using iterative inverse methods. The fact that the alternative method (perfect plasticity) does not require velocity information seems irrelevant to us. In case class 3 would not exist as a separate class, most surging glaciers would fall in class 2 and would hence be modeled with PISM instead, in which case the inversion would rely heavily on the velocity data to invert for friction. This would create major errors in the thickness inversion when velocities are strongly overestimated (as is the case when a surge happened in 2017-2018 while the DEM is from ~2011 and the dh/dt dataset from ~2015). Arguably we could alternatively have selected surging glaciers based on the timing of the DEM (2009-2012) or dh/dt (2010-2019) datasets instead. However, since the Koch et al. (2023) dataset does not go back further than September 2015, we would have trouble selecting all glaciers that surged during the periods of the dh/dt and DEM datasets. Finally, we would like to note that all 13 glaciers that Koch et al. (2023) listed as actively surging between late 2015 and 2018 were treated as surging glaciers in our approach and were hence modeled with the perfect plasticity assumption. We in fact mistakenly classified Nathorstbreen as surging, which is the only glacier that surged in 2015-2016 and not in 2017-2018. To be correct, we now state that all glaciers that according to Koch et al. (2023) surged between 2015 and 2018 are treated as surging glaciers and included in class 3.

The suitability of the perfect-plasticity method for surging glaciers is hard to verify. We do know from tests that the iterative inverse methods (using PISM and IGM) would introduce major outliers in the thickness map. This was the initial reason to search for an alternative method for these glaciers. A benefit of using the perfect plasticity method is that (in our case) it uses a DEM that is from before the surges initiated, which makes that the thickness inversion corresponds to the quiescent phase. We think this is an advantage because the strongly transient physics and stresses involved in an active surge may not be well captured by the perfect plasticity assumption (or any other ice flow model). We add the following in Sect. 3.3:
"*In the perfect plasticity assumption ice thickness is controlled primarily by the surface height (Eq. 1). Since the DEM (2009-2012) was collected prior to the initiation of the surge for the selected glaciers, the thickness estimation is effectively based on the pre-surge glacier geometry. We regard this as an advantage as ice flow models in general are not well able to describe the strongly transient stress-state of actively surging glaciers.*"

MINOR COMMENTS
L126 'This comment was undressed (neither in the rebuttal nor in the revised manuscript).' From my understanding, the term yield stress relates to when deformation becomes possible whereas the term sliding law normally relates basal velocities to general basal stress conditions. Please check this terminology and be specific about any assumptions.

Good point. Indeed the term yield stress refers to the basal stress required for sliding to occur. This applies to the plastic Coulomb sliding model. Since we rather use a linear sliding law in PISM, the term yield stress may be a less appropriate term to use. We hence change this and now use the term sliding coefficient (C) instead when referring to the sliding law used in PISM. We still use the term "yield constant" in the description of the perfect plasticity assumption, where we think the terminology is appropriate.

L270 'Follow-up on your answer to:' For the land- and marine terminating glaciers, you report mean thickness values of 42 m and 162 m, respectively. How can you reconcile this value with the archipelago-wide average of 205 m. I certainly miss something here ;o) Obviously, I missed the difference between median and mean values. This metric-mix makes it very hard to compare. Solutions would be to either use both metrics for each time you specifically provide averaging measures or only give either of the two metrics consistently throughout the manuscript. Personally, I prefer the former. Your decision.

The large thickness difference is not because of the use of median or mean, it is rather because different quantities are compared. The mean thickness of 205 m is calculated by dividing the total glacier volume by the total glacier area. The median thicknesses for land-terminating glaciers (42 m) and tidewater glaciers (162 m) are the median values of average glacier thicknesses (per glacier) for all 1363 land-terminating glaciers and 181 tidewater glaciers, respectively. In other words, the median values are found by sorting the glacier-average thicknesses for the glaciers in one class from low to high, and then taking the midpoint value (682nd value for land-terminating glaciers; 91st value for tidewater glaciers). These median values are hence strongly affected by the size distribution of glaciers in the RGI dataset. The relatively high number of small (thin) glaciers in both classes explains the discrepancy between the Svalbard wide mean thickness and the median thickness per glacier class. Anyway, we understand the confusion and have added the following to Sect. 4.1:

"*These median values are much lower than the Svalbard-wide mean ice thickness (205 m), which results in a skewed size-distribution with predominantly small and thin glaciers in both glacier categories (LT and TW)*."

and to the caption of Fig. 5:

"*The box-plots for LT and TW glaciers are based on mean thickness values for every glacier.*"