Review of van Pelt and Frank (2024): A new glacier thickness and bed map for Svalbard

Summary
This paper presents a new ice thickness and bed map for Svalbard based on a method previously developed by the authors. Here, they use three different methods for small, land-terminating glaciers (IGM), larger and tidewater glaciers (PISM), and surging glaciers (perfect plasticity) to derive their results, having performed a considerable amount of calibration and processing to reduce errors and ensure agreement across their results (no big jumps in ice thickness at ice divides). They then compare their results to other recent bed-thickness datasets for Svalbard, showing that their work sits within the expected range, but substantially reduces errors and bias across the board.

I think this is an innovative paper that attempts to leverage recent developments to obtain the best-possible results. However, I have a few major concerns, as well as several minor ones before I would be happy to recommend the paper for publication. I wonder whether the use of three different methods, particularly when it is not clear to me the rationale for using two different ice-flow models, has not overcomplicated the paper and sacrificed internal consistency – normally an advantage of these large-scale datasets – for an unclear gain in accuracy. The authors establish using PISM where they use IGM would lead to a worse outcome, but not whether the reverse is true, which seems to me a major oversight that makes it difficult to see what advantage using PISM and complexifying the method really brings. I am also very unclear as to how the authors set up their model domains and how they dealt with the resulting boundary conditions, which makes it difficult for me to assess the quality of their modelling. Overall, I therefore think major revisions are required: it may just be a question of adding/clarifying some information that is not obvious as the paper is currently written, which I hope is the case, as I think the end outcome and method are very interesting!

I should also note that I read the paper and wrote the review before I read Reviewer 1's comments. The fact that both of us largely raise the same issues is therefore not a result of groupthink.

Line and page numbers refer to those in the manuscript.

Major Points

- Consistency: this is perhaps something of a more philosophical point, but a major advantage of these kinds of large-scale studies is that (usually) they apply the same processing steps to a wide area so that, even if one doesn't believe the absolute numbers very much, one can be confident that the results are internally self-consistent. Here, the use of three (very) different methods means this cannot be taken for granted in the same way. I think the authors have done a lot of work to try to overcome this, particularly with interpolating results onto different grids, but I wonder if this was the right approach.
- Choice of methods: I do not think the authors really provide a clear justification for why they chose to use PISM for the larger (and tidewater) glaciers, as opposed to IGM. I take the point that PISM in SIA+SSA mode does a good job for larger and tidewater glaciers, but IGM is a higher-order model that would also do a good job here. Usually, this choice would be justified by saying that higher-order models are too computationally expensive to make them practical at this scale, but IGM has been explicitly designed to run fast and overcome that objection. So what motivated the choice? Because, if the authors had used IGM for all the larger and tidewater glaciers, that would have allowed them to achieve a much larger degree of consistency in the results and avoided them a considerable degree of work at the calibration, method and post-processing stage, it seems to me (of course, they would have had to find a strategy to vary the sliding parameter in IGM, but that feels to me an easier thing to do than have two separate methods working on different assumptions at different resolution)
- Boundary conditions and model domains: I am unclear how the authors defined their model domains. I assume, given they use RGI6.0, that they take each RGI outline and invert it individually? Or do they take all contiguous RGI6.0 outlines and invert them as one entity? In the former case, how do they then deal with ice-ice boundaries, where two different RGI entities are in contact? In both cases, what boundary condition is imposed at the front of tidewater glaciers? As both the PISM

and IGM methods involve small forward timesteps, these issues need to be considered. At the very least, a few lines in the discussion about how not considering these likely introduces some local inaccuracies need to be added.

- Discussion: Ultimately, I think this comes back to my point on the choice of methods above, but I don't find that the discussion does a very good job of highlighting what this study brings to the table and why people should use the bed calculated in this study as opposed to those from other studies. The authors provide plenty of description for how their results compare to other datasets, but mostly do not analyse why these differences occur, making it hard for readers to assess which product is better for their particular application. The fact that it is also unclear as to why the authors made particular methodological choices (see my comment above) then further muddies the waters here. The authors do show that they substantially reduce the error on larger glaciers compared to previous studies and the bias across all glaciers (Table 2), which I would argue is the main selling point of their results in the current formulation of the paper, but this gets a bit lost in the discussion and no mention of it is made in the abstract (there is a partial reference in the conclusion, but only to the error on larger glaciers), making it very easy for readers to lose sight of it completely.

Minor Points

- p.1, l.10: I might venture to say that a mean ice thickness at the scale of the whole Svalbard archipelago isn't that useful or meaningful a number to include in the abstract (at least, as a headline figure for the paper, it seems some way down the list of things that readers would want to know)? The total volume, yes, but I would suggest maybe converting that into an SLR equivalent for the second number, or reporting the maximum ice thickness, which is something that makes a bit more sense at that scale. Or, possibly even more useful, say something about how the volume estimate presented here compares to other studies' estimates.
- p.2, l.42: Reference formatting for Farinotti et al.
- Table 1: Why use the 20 m NPI DEM when it has to be downscaled to 100 or 500 m immediately? Wouldn't the COP90 DEM have been a better choice to fit with the modelling resolutions and also sit more in the middle of the range of most of the other data (2010-2019ish)? Also, the RGI6.0 outlines for Svalbard have dates of 2000-2010, so please update the table to reflect that.
- Figure 1: I can see already that, on Austfonna, there are two methods being used to generate the results, despite the assertion in lines 116-117 that all the glaciers connected to larger ice caps are modelled using PISM (i.e. the same method). Can the authors confirm whether the figure or the text is correct here? If the figure is right, how are they dealing with the jumps in thickness at the ice divides on Austfonna?
- p.7, l.169: 'with a'
- p.9, l.200-204: Can the authors comment as to how far using uniform parameter values might introduce some error into the results?
- p.9, l.214: I confess I'm not entirely clear on why the thickness field produced by IGM would have gaps in it that need interpolating?
- p.10, l.231-232: This seems to me quite a substantial upsampling of the majority of the dataset that might introduce a considerable number of artefacts. Would not downsampling the 100 m proportion of the dataset to 500 m have been the more conservative choice?
- p.10, l.245-257: Yes, but are there observed glaciers in each of the three categories, such that all three types of inversion are bias-free? More generally, with three different inversion methods, would there not need to be three separate estimates of sigma H bar, one for each category? Because a mean error of 3.5 m on ice thickness across the whole of Svalbard seems a little too good to be true. The observations themselves would have bigger errors than that!
- p.16, l.314-319: Have the authors performed the same comparison in the other direction? As in, what happens if IGM is used to model the larger glaciers where PISM was the preferred method? Otherwise, I think it's difficult to say that the combination of the two methods is superior to either alone. The approximations in PISM may be suitable on the larger glaciers, but it doesn't follow that that means they're more suitable than using a higher-order model
- p.16, l.321: OK, yes, 6855 is higher than 6800 and 207 is higher than 205, but I'm not sure that it's really a meaningful difference, especially when both those numbers are well within this study's own error bars. Consider rephrasing this to make it clearer that this study's integrated volume and mean thickness results are not significantly different to those from Millan et al.

- p.16, l.320-333: Could the authors provide some more analysis of why these differences exist? They posit sensible reasons for why Millan et al. likely overestimate ice thickness on larger glaciers, but I think it would make the paper much more useful for the community if they can suggest some reasons for the other differences (spatial distribution more similar to Fürst, thicker ice at lower elevations than Fürst, less pronounced jumps at ice divides than Millan, etc.), as it would help people work out which is the best bed product for them to use for their particular application
- p.18, l.384: This is only true provided other people use the exact same set of final modelled bed, velocity, surface, etc. as used in this study as their initial conditions. If someone took the bed from this study and then used, say, the COP90 surface DEM and ITSLive velocities to initialise their model, they would not have a harmonious set of initial conditions. Please rephrase this to make it more clear.