Comments on “Regularization and L-curves in ice sheet inverse models, a case study in the Filchner-Ronne catchment”

General comments

Wolovick et al. proposes best-practices for achieving a well-balanced regularization in the context of basal drag inversion. Inversions are commonly used to initialize (and calibrate boundary conditions on) ice sheet models before running numerical experiments (e.g., transient simulations). Inversions based on minimizing a cost function (as is the case of this manuscript) is an ill-posed problem. Adding some regularization tends to make the minimization problem less ill-posed and converge to a global minimum. The type of regularization is somewhat subjective. The commonly employed one is the Tikhonov regularization of first order (as used in this manuscript), but many others have been proposed. The weight of the regularization term itself may be adjusted depending on the level of smoothness that the user wants to enforce. One way to determine this level of smoothness is to perform an L-curve analysis, which is frequently used in the glaciological community. In this paper, the authors performed several numerical experiments (inversions) to infer the friction coefficient varying some modeling choices, such as the mesh resolution, type of friction law, stress balance model etc.

Overall, the paper is well-written, the figures are well done and their captions are clear. However, the description of the numerical experiments could be written right in the introduction - it is hard to follow what exactly was done and why before digging into the methodology and results sections. Also, there are unnecessary technical details, discussions and speculations in several sections, which makes the text unnecessarily long. Most of them could be removed or moved to a supporting information file.

The biggest issue with the manuscript is the authors claiming they have “the best” recommendation for performing the friction coefficient inversion (by minimization). I would argue that it is simply impossible to validate any inversion without any observation of what we are trying to invert for. How can one claim that an inversion is better if there is no information about the “real” friction coefficient or basal drag? Affirming that they have the “best” inversion results (or knowledgeable friction coefficient, or Best Friction Map) only based on the L-curve shape alone is problematic. One could change the norm of the cost function (e.g., L1 instead of L2 - the norm used by the authors), the regularization term, the descent algorithm, line-search scheme, the inversion method, and find completely different recommendations based on the shape of the L-curve (with or not minimal values of cost functions), whatever “better” means here. Also, minimization problems have lots of local minima; so if the descent algorithm constantly achieves global minima, the results are biased, and not the “best”.

There is not much to say about the methodology itself without comparing its results with a known friction coefficient and other methodologies. It would be better if the authors employed a simpler ice sheet domain, with a flat bed, and a ***known*** friction coefficient (e.g., based on the ISMIP-HOM setups for example) to compare with their inversion results. The resolution of BedMachine is 450 m. It means that any element of size $h > 450$ m (considering linear finite elements) will not catch the complete geometry (i.e., bed elevation), which adds another factor to the current results. Changing mesh resolution also changes geometry (bed), so you are not solving the same mathematical problem (for any $h > 450$ m). Also, it is hard to say the regularization parameter λ goes to a nonzero number based on the extrapolation of a fitted curve (Figure 8b). Is there any guarantee of that? Could one have similar results if using another inversion scheme, descent algorithm, regularization term, cost function norm (L1, L2, Linf), cost function (log vel) or numerical setup? Additionally, affirming that the exponent m (in Weertman's friction law) should be >1 based on the shape of L-curve alone is simply impossible: can we really select the correct physics without direct observations of that, based only on one kind of ill-posed inversion methodology? A reasonable approach (but not the best) is matching the complexity of the model with the accuracy of the observed data (Smyl and Liu, 2019). But is it possible to be tested in a glaciological context considering the current lack of observed data below ice sheets?

With a simpler control model (synthetic model with a simple geometry and known friction coefficient) one could isolate and compare the effects of different practices and methodologies, regularization terms, cost function scaling, mesh resolution, errors in observed data, friction law, different wavelengths of friction coefficient, stress balance, ice temperature, etc, and then point out which approach achieves a reasonable result considering the known friction coefficient (see for example Smyl and Liu, 2019).

I do agree with the “fitting the data using the least amount of structure” message of the paper, but is it not already known by the glaciological community?

With that said, and based on the TC criteria (novelty, rigour, impact, presentation quality) it is not possible to support publication. However, I do feel the manuscript, after several changes to the text, removing unnecessary discussions and speculations and with focus on a single message, could be a good educational paper for beginners in ice sheet modeling (I am not sure if TC is the right journal). It brings some interesting approaches that could (but not necessarily guarantee) improve the way a corner in L-curve could be estimated, shows the impact of mesh resolution on the inversion (considering that bed elevation changes as practical mesh resolution changes), compares nonlinearly in friction laws, etc, which could help others to understand and run inversion problems in glaciology.

Specific comments

lines 8-12: For the abstract, I believe it should be a little bit clearer, mainly for a broad audience. The results or conclusions are written without any mention of the methodology - for instance, a hydrological model was used? So one can not catch the main results without reading the methodology part in the paper. A suggestion is to briefly describe the methodology, and then briefly take home messages. A shorter abstract in general is more readable than a long one.

line 13 velocity data. Do you mean ice surface velocities?

line 14 new constraints. Do you have any examples of possible constraints?

lines 29-30. Not sure if these studies show temporal variability of the bed. Do you mean basal drag or bed elevation?

line 34. continuous problem?

line 35-36. Note that this is true if the discretization of the basal drag is P1 over a mesh, for example. But one can model the basal drag using any other polynomial order, even piecewise constant. Maybe changes to “the number of free parameters is finite and depends on the discretization method (for example, in a grid….)”, or something like that.

line 40-44 I believe this part “Regularization involves adding .. minimized in the inversion.” could be shortened. There is “adding” and “combining”, which basically is the same.

line 51. What is the meaning of non-convergent here? There is no an L-shape or corner?

line 52: solutions: solution of the PDE, or the inversion process (inverted field)?

line 52-53 the phrase “specifically, L-curve analyses are … Fourier coefficients of the solution” is not clear. For example, which “operator”? Is it performed a Fourier analysis? If the technical detail here does not add relevant information, maybe remove it. Stick to the point that in some problems (which ones?) the L-curve does not converge (what means converge?).

lines 56-65 The paragraph is what is expected to be (briefly) written in the abstract (please, see my comment about missing a few phrases about the methodology), so the reader can at a glance know what was done and the resulting conclusions.

line 80 “without the need to derive the adjoint”. Maybe it is worth mentioning the meaning of adjoint here (some readers may not be familiar with).

line 81-82 “they did not rely on a series expansion representation of the drag coefficient in their derivation”. Maybe it is worth adding after that phrase how the authors discretized the basal drag.

lines 90-99 I think that paragraph should be shortened. There is a discussion that is beyond the scope of “review”. There is not need to say that one group did something in the second paper and not in the first. My suggestion is to stick to what is relevant of the current manuscript: there is not much detail of the regularization? Does the choice of regularization impact the basal drag near the grounding line? (note that the meaning of “grounding line” was not mentioned before - it is important for readers not familiar with specific terms of the cryosphere community);

lines 100-104. The same – the paragraph should be shortened.

lines 109-121. The same – the paragraph should be shortened.

line 138. The references “Rignot et al., 2017a” and “Rignot et al., 2017b” are the same.

line 150-151. Maybe it is worth mentioning that the operations (\nabla) are on the map-view plane. The same for u = ux, uy

line 164. There is no detailed mention about how the hydrology model is used in the inversions. It is used to compute the effective pressure? Maybe it is worth mentioning that a hydrological model is employed in the last paragraph of the introduction.

line 168. “of our model”. Which one? Thermal model?

line 182. “by our model” => by the SSA model (or something similar)

line 182. “varies by about” => varies spatially (or on map view plane) by about

Figure 3c. What is hydraulic diameter (x-axis label of panel c)

line 194. “of the mask”. Which one?

lines 195- 199. Maybe it is worth mentioning that err, hmin and hmax are Bamg parameters.

line 201. “anisotropic elements”. Which anisotropy ratio? Note that anisotropic elements might introduce numerical errors.

line 202. “common diameter”. What do you mean?

Line 203-205. “We analyze fast-flowing …to generate the weighting function”. Maybe it is better to invert those phrases: explain first how you compute the weighting function, then the nominal cutoff.

line 206-2067. I am not sure what it means: “and also separating reliable velocity data from unreliable data”. If it is not relevant, maybe cut that phrase.

line 214-215. “This technique is superior to simple interpolation”. Well, I am not sure if it is superior, since you are performing more operations than a simple interpolation and smooth things around. Maybe just mention that “this technique prevents aliasing …”. Also, as said above, most of the details here could be moved to a supporting information file.

line 217-220 Does this smoothing procedure produce artifacts in terms of grounding line position? Any change (or parameterization) of the grounding line impacts the stress balance.

line 231-232 Equations (5) and (6). “B” here is the bed elevation, right? Note that you use the same letter for the rheological parameter.

lines 286-293. It is not clear how the S_{reg} is obtained.

line 306. “inverse model to convergence”. Which descent algorithm did you employ? LBFGS (M1QN3)? Which parameters to stop? Maybe it is worth mentioning it in the text.
line 309-324. I think most of the details here could be moved to supporting information. Maybe just say you applied a fitting curve to the sample points and the reason you did that (needs of a smooth 2nd derivative etc).

line 327. “curvature drops to 1/2 of its maximum value”. Is 1/2 a heuristic value?

line 329-330 I think the phrase “and the equivalent uncertainty ratio on a linear scale is simply the exponential of that uncertainty” could be deleted.

lines 332-343. Entire paragraph. This is what I expected to read somewhere above, such that the reader could have a big picture of all the experiments you did. Not sure if moving it or writing something similar above would improve the manuscript. Nevertheless, my suggestion is to keep it as simple and clearer as possible, with few words.

line 350 “giving us confidence in our inversion procedure”. I am note sure if achieving L-curve monotonicity will give you a confidence of the inversion itself (i.e., achieving global minima); there are other factors involved (eg. descent algorithm convergence criterion). What the monotonicity could say is that your algorithm is not introducing potential spurious or numerical errors, as you mentioned below. Maybe remove or rephrase that phrase.

line 362-363 - is it not the contrary?

Line 364-369 – “The power-law nature… that includes both limbs”. I am not sure if this discussion should be here. Maybe remove it.

line 385-386. I believe it is important to mention you also performed a spatial and spectral analysis of the inversion results somewhere above (maybe in the introduction).

line 447-448. “These results emphasize the importance of choosing the correct regularization before using an inversion to draw conclusions about the short-wavelength structure in the ice sheet basal drag”. I believe this phrase could be removed.

line 480. “Once mesh resolution is better than the ice thickness” better you mean “finer”? Maybe it is worth rephrasing it.

line 504. lambda_best, lambda_min (is it the opposite? lambda_best. lambda_max) decreasing resolution, increasing lambda

line 515 What exactly do you by “no geological variation in the substrate”?

line 624 Best Drag Map. Combining several inversions to produce a single basal drag map seems a reasonable approach; however, I would not say “best” map. Unfortunately, we do not have measures of the “real” basal drag to compare with. Better say a combined map (an averaged or mean map) that includes some modeling “uncertainties” or different physical parameters (note that uncertainties in data sets were not included - e.g., bed elevation, surface elevation, ice velocity etc.)

line 632-633 “Each inversion is weighted according to the inverse of its total variance ratio.” Is there a reason to use that weight?

line 663 “singular values of the operator”. Which operator?

line 675 by a mesh of that size => by a mesh of that resolution
As mesh size approaches zero => As element size approaches zero or As mesh resolution approaches infinity (or anything better)

line 726. “future modelers have no excuse for continuing to use linear Weertman sliding laws.” It sounds like the whole community is using m=1 nowadays. I am not sure if this is really true, although you mentioned some modelers employed m=1 in the ISMIP6 effort. The most common is m=3 and new studies say that m>3 seems more appropriate. I would rephrase it appropriately.

line 731-732 - I believe the delimiters “—” in that phrase is misplaced. Please check it.

line 727-745. Note that in unconfined ice sheets, any change in the ice shelf will not propagate inland (Schoof, 2007). Probably the retreat of the grounding line is caused by loss of buttressing (which is caused by shelf thinning due to increased basal melting).
Also, the phrase: “the nonlinearity of the sliding law also has a strong influence on our predictions of future ice sheet dynamics”. From Barnes and Gudmundsson (2022): “In broad terms, the system always reacts in the same way to a particular change in forcing, no matter which sliding law or parameter values are used. The magnitude and speed of these reactions vary, but the response is clearly bounded.” That said, I suggest rephrasing or removing that discussion, which I think is unnecessary here.

line 785-800. Again, not sure if this whole discussion is really needed. Better shortening or removing it.

Technical corrections

line 62 vertically integrated model => vertically integrated ice sheet model

line 63 our various model results => maybe “our various inversion results“

line 84 Rather than introduce => Rather than introducing

line 116. components very inversely => components vary inversely

line 139. driving stress resolved in the flow direction => driving stress projected on the flow direction

line 149. check the equation

line 163 HO models => HO model

line 207. with mesh number => in all meshes (maybe?)

line 208. the model places => the mesh generator (or mesher or Bamg or our meshing strategy) places

line 212. all relevant grids => all relevant fields (I know the fields are defined on grids, but better saying fields (defined over equally spaced grids))

line 224. please, un bold subscript “b” in “\tau_b”

line 264 analyze structure in C => analyze the structure in C

line 265 under-penalize structure => under-penalize C (maybe?)

line 281. We get this guess => We compute this guess

line 284. driving stress resolved in the flow direction => driving stress projected on the flow direction

line 387. we tested differed by => we tested differ by

line 390: please, check if “\tau” should be bold here (same in lines 392, 411, and elsewhere)

line 431 driving stress and basal drag are locally balanced => driving stress and basal drag are virtually locally balanced (not sure if coherence in power spectra means exactly \tau_d = \tau_b)

line 467 represent mesh size => represent element size (and elsewhere in this subsection)

line 483 we can get an approximate = > we can estimate an approximate

Figure 8. “with solid diamonds marking λbest for each mesh.” it is hard to see the diamonds. Maybe changing their colors to a single color? (e.g., black) (same for Figures 10, 11 and 13)

Table 1. “C_0” should it be C_W?

line 519. geometry-based fields => geometry-based N fields. Same for “all three fields” => all three N fields

line 559 Fig. 10a => Fig. 11a

line 560 Fig. 10c,e,g => Fig. 11c,e,g

line 625 descriptions of the ice sheet bed => descriptions of the ice sheet basal drag (or friction coefficient)

line 633. we use the version of total variance => we use the value of total variance

line 708. coulomb => Coulomb (and elsewhere)

line 962 mesh size => mesh resolution

References

Barnes, J. M. and Gudmundsson, G. H.: The predictive power of ice sheet models and the regional sensitivity of ice loss to basal sliding parameterisations: a case study of Pine Island and Thwaites glaciers, West Antarctica, *The Cryosphere,* 16, 4291–4304, https://doi.org/10.5194/tc-16-4291-2022, 2022.

Schoof, C. (2007), Ice sheet grounding line dynamics: Steady states, stability, and hysteresis, *J. Geophys. Res.*, 112, F03S28, doi:10.1029/2006JF000664.

Smyl, D., Liu, D. Less is often more: Applied inverse problems using hp-forward models, Journal of Computational Physics 399 (2019) 108949. https://www.sciencedirect.com/science/article/pii/S0021999119306540. doi:https://doi.org/10.1016/j.jcp.2019.108949.