General Comments

Wolovick et al present a detailed study centred on the regularization of an ill-posed problem in glaciology, namely the estimation of a basal traction field from surface velocity data. As the authors say, many ice sheet modellers rely on these techniques and improved discussion is welcome. I approve of the choice to confront real-world data rather than only recover pre-defined fields. The numerical experiments are sound and the analysis thoughtful. The demonstrations of scaling the cost function are particularly useful.

The paper is quite long: I don't think much is added by the inclusion of Budd sliding experiments, which are not the pressure dependent laws of interest today. I recommend removing these sections (they could go in a supplement?)

This is a good paper and should be published on the basis of its results and the detailed exploration of interesting aspects of the inverse problem, but sometimes seems to insist that its methods are optimal without rigorous proof, while using language that seems quite acerbic when referring to the work of others. I think the authors should word their points more carefully. I don't think that will require much work or change the paper in a major scientific sense.

For example, starting in L680, "It is common in the inverse modeling literature to read some version of the **sentiment** [I object to this word] that we … cannot use inverse models to distinguish between different sliding laws.", and says that the "the purpose of inverse modeling,… is not merely to fit the data, but to fit the data using the least amount of structure". In my view Joughin is correct to say that even complete knowledge of *Tb* alone at a single time provides no information on the relationship between *Tb* and *u:* **it is not a sentiment and does not fundamentally misunderstand anything.** Note that Joughin 2010 and Joughin 2019 make use of data at multiple times to determine that *m* is not 1, which you should cite in the paragraph around line 710. I do agree that having more structure in C than Tb is undesirable/unjustified, but it might still be the underlying truth. You are then claiming that the regularization makes more information available, but Tikhonov regularization is a reasonable bias towards coarser structure rather than anything fundamental or inevitable. At the same time, having finer resolution observations could permit more structure (in Tb, not C) to be determined and that could be desirable. In other words, fitting the data using the least structure sounds a decent aim given the ill-posed nature of the problem (but what does it mean 'to fit', when the misfit is suboptimal?), but equally, one could claim to seek the most structure given the information content of the data (accept a suboptimal misfit in the light of knowledge that the problem is ill-posed so that certain types of observational error will be amplified and so certain aspects of the solution will be dominated by error).

Specific Comments

Abstract.

L-curve analysis is not the only way to select the regularization parameter, and so it cannot be used to select the optimal regularization level in general. It can be used to the select the optimal parameter if one accepts the idea behind the L-curve, but an alternative approach, which has been explored at least once in glaciology (Martin and Monnier, 2013), is Morozov's discrepancy principle.

I'm not saying you need to review this in the abstract but rephrase to be clear that L curve analysis is heuristic (but as you say, should be done properly).

Review

L 85. Stopping an iterative optimization method 'early' can be a type of regularization, depending on the iterative method of course. See e.g Hansen 1994.("the CG process has some inherent regularization effect where the number of iterations plays the role of the regularization parameter")

Methods

L148: Eq 1 is wrong – but I see the authors already note this in a TC comment.

L162 (3.3): This is reasonable way to estimate 'B-bar' (and I also find that you have to reduce shelf temperatures by about 10 C, even with advection, unless I take care with the bottom boundary condition in the shelf). However, I think you should note that the full inverse problem includes estimation of 'B-bar' because Glen's flow law depends on unknown thermal conditions and even then is not the whole of large-scale ice rheology (e.g damage in shear margins might be important) This is obviously a problem because it makes the inverse problem even more ill-posed (i.e underdetermined as well as ill-conditioned) , and so assumptions such as the one made here are needed.

L212. Is this method entirely invented by the authors or is there a reference? It seems strange to me to call velocity a 'grid', when it is 'data on a grid'. At any rate, I can't see from what you say how the technique works.

314. This paragraph can be confusing at first, because 'smoothing' can apply to both 'C(x,y)' and the L-curve. Perhaps introduce $J_{r,c}$ andJ_d, c first.

L325. I like that you explore the region around the optimal λ but the bounds you choose are arbitrary and so should not be called 'minimum/maximum acceptable', and you don't 'bracket' the full range of the corner.

Results

L365 : A very interesting point.

L370: Why does the curvature of the inidividual components matter? This is a long paper, so perhaps this detail could be trimmed.

L450. These are good points, but I am not sure that the points about Vogel 1996 help argue them (Vogel calls a method convergent if the regularized solution tends to the exact solution as either the upper bound of the magnitude on error tends to zero, or the number of observations of a random 'white noise' variable tends to infinity, model node count is secondary). The real point is that the 'forward' models are approximations that neglect fine structure.

L470 'but in addition coarse meshes are also unable to fit the data as well as fine meshes' or indeed, approximate solutions solve the stress balance equations.

L478: 'e-folding mesh size': define.

L500 (while paragraph). This is not convincing to me: what is meant by approximate bounds?

L555- (nonlinear sliding section). This is for me the most interesting result. What happens if you take your m=1 coefficient fields ($C_1(\lambda)$) use the formula $C_3|u|^{\frac{1}{3}} = C_1|u|$ to work out the corresponding fields for $C_3(\lambda)$ and then compute $J_{reg}(\lambda)$? Do the points lie on top of the m = 3 L curve (so that the difference is all in the cost function analysis) or not (so there is an effect in the individual problems, presumably does to the way the regularization term affects the optimization)

L627; "We feel that it is more appropriate to produce a consensus view of drag, τb". Agreed.

Discussion

L862-905 – this is a long section for what is at the end of the day opinion/ speculation. I don't think it is wrong, just not part of the work that has been done here. Same regarding 970 onwards.

References

Martin, Nathan & Monnier, Jerome. (2013). Of the gradient accuracy in Full-Stokes ice flow model: basal slipperiness inference. The Cryosphere Discussions. 7. 3853-3897. 10.5194/tcd-7-3853-2013.

Hansen, P.C., 2007. Regularization tools version 4.0 for Matlab 7.3. *Numerical algorithms*, *46*, pp.189-194.