Response to Reviewer 2

Original reviewer comments are in black, our responses are in blue.

We would like to thank both of the reviewers for their thorough and thoughtful reviews. The tone and recommendations of the two reviews were quite different, with Reviewer 1 recommending rejection and Reviewer 2 recommending acceptance with minor revisions. In addition to disagreeing on their overall recommendations for publication, there were also several instances in which the reviewers gave conflicting recommendations for specific changes or revisions. For instance, Reviewer 1 recommended that we should have used synthetic examples for our inversion, while Reviewer 2 praised our decision to use real data. Nonetheless, we have attempted to construct two replies that harmonize the various recommendations, and we have also given our rebuttal to Reviewer 1's reasons for recommending against publication.

While the official instructions we received explicitly called for us not to prepare a revised manuscript at this stage, we found that it was easier to deconflict our responses to the two reviews through reference to a common revised text. As a result, our revised manuscript is already partially finished, and the remaining requested changes should be relatively minor. Not counting smaller wording changes and other minor corrections, the biggest changes to the new manuscript are:

1. In order to shorten the manuscript, we removed the paragraphs speculating on future 3D inversions (lines 868-906), the paragraph speculating on future coupled hydrology/drag inversions (lines 785-800), and the final two paragraphs of the discussion about nonlinear sliding (lines 716-745).
2. In order to further shorten the manuscript, we moved several parts of the methods from the main text to a supplemental section. In particular, the methods sections on thermal structure and rheology (Section 3.3), mesh generation (Section 3.4) and the description of our curve-fitting procedure (Section 3.7) have all been moved from the main text to a supplemental methods section.
3. In response to comments from both reviewers, we moved the section explaining our experimental design (Section 3.8) from the end of the Methods section to the beginning.
4. In response to comments from both reviewers, we will add a supplemental section explaining our multi-wavelength smoothing procedure for interpolating gridded data onto our mesh.
5. In response to a comment from Reviewer 1, we will add a second panel to Figure 14 showing the uncertainty about our best combined drag estimate. We have already computed this uncertainty field and it is included in our data release on Zenodo, but we did not visualize the uncertainty field in the first draft of the manuscript.
6. In response to a comment from Reviewer 1, we will add a supplemental figure showing the distribution of mesh element anisotropy.
7. In response to a comment from Reviewer 2, we will add a supplemental section showing the results of a test where we convert the drag coefficient from the linear Weertman inversion into a nonlinear sliding law, and then compare the resulting L-curve with the L-curve obtained for the nonlinear sliding law itself. We already performed preliminary

General Comments

Wolovick et al present a detailed study centered 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?)

While we agree that the manuscript is quite long and needs to be shortened, it is not necessarily true that Budd laws are no longer of interest in the glaciological community. The Budd sliding law continues to be used in recent glaciological publications (e.g. Kazmierczak et al., 2022; Brondex et al., 2019; Barnes and Gudmundsson, 2022). In addition, at least two of the models in ISMIP6 used Budd sliding laws (Seroussi et al., 2020). Furthermore, the use of a Budd sliding law in the context of our experiments makes sense, as a Budd law is easily comparable to a Weertman law with the same stress exponent. By using both Budd and Weertman sliding laws, we can set up experiments to directly test the influence of effective pressure with everything else in the inversion held constant.

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.

Thank you. We agree that our wording was too harsh in places. We will rewrite the relevant parts in the revised manuscript.

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.

Thank you for pointing out those references. We will add these citations to the paragraph beginning on line 707, and we will soften our criticism of Joughin (who has made many important contributions to the field of glaciological inverse modeling over the past two decades), as well as dropping the word "sentiment".

However, we hold to our assertion that inversion results at a single snapshot in time can be used to learn something about the form of the sliding law. As we discussed in our reply to Reviewer 1, Occam's Razor suggests that, when two competing parameterizations obtain a similar fit to the data, we should prefer the parameterization that requires less complexity to do so.

Of course, this is not definitive proof that the sliding law has any one particular form. As you rightly point out, it could be the underlying truth that slip and drag have a linear relationship, and the extreme dynamic range of C could just be the unfortunate reality. However, in order to conclude that this is so, we would need positive evidence from other sources that a linear relationship between slip and drag is the correct one. The inversion results themselves may not be definitive proof of nonlinear sliding, but they are not neutral between all sliding laws either. If you accept Occam's Razor as a useful organizing principle for guiding scientific inference, then the inversion results (even for a single time slice) are evidence in favor of nonlinear sliding. That evidence could certainly be overridden if other lines of evidence favored linear sliding, but, as we discussed on lines 707-715, other lines of evidence agree with the inversion results in favoring nonlinear sliding. The inversion results add to a convergence of multiple lines of evidence favoring nonlinear sliding laws. We will add the two references that you suggest to this discussion.

As promised in our reply to Reviewer 1, we will soften our wording in this section to make clear that inversion results are supporting evidence in favor of nonlinear sliding, not definitive proof. However, we do disagree with your (and Joughin's) statement that "even complete knowledge of Tb alone at a single time provides no information on the relationship between Tb and u". Our point here is that competing parameterizations can be judged on more than just the quality of their observational fit, they can also be judged on the complexity of the free parameter(s) needed to obtain that fit. The new advance we are proposing here is that the complexity of the inverted free parameter required to fit the data provides additional information that can be used to discriminate between competing parameterizations. It may not be definitive proof, but it is not "no information" either.

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).

This is a good point. Lambda_best provides the best trade-off between observational fit and inverted complexity according to L-curve analysis; this could be thought of as finding the minimum necessary structure to explain the observations. However, a modeler could reasonably choose to use a smaller value of lambda if they valued obtaining the highest possible resolution of the basal drag more than they valued minimizing complexity in the inverted field. Alternatively, a modeler could reasonably choose to use a larger value of lambda if they had little interest in detailed basal structure and only wanted to include the structure that is absolutely required to produce an approximately reasonable velocity field.

This is why we bracketed the range of acceptable lambda values with lambda_min and lambda_max, rather than giving lambda_best alone. As both reviewers pointed out, bracketing the range of acceptable lambda in this way involves the selection of an arbitrary "heuristic" threshold for L-curve curvature (we used a value of ½ the maximum curvature to define lambda_min and lambda_max). Nonetheless, the drop-off in total curvature below lambda_best is quite steep, so this threshold has little influence on selecting lambda_min (it has a bigger effect on the selection of lambda_max, which we will explicitly mention in the revision). Thus, modelers who are interested in obtaining the highest resolution picture possible can still use this method, and the L-curve method will give them confidence that lambda_min will allow them to find the highest-resolution picture of the ice sheet basal drag that can reliably be obtained from the given observations.

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).

Reviewer 1 also mentioned alternate choices that we could have made in quantifying misfit and structure, such as using the L1 norm instead of the L2 norm. The L2 norm is by far the most common in glaciological inversions, and L-curve analysis is the most common method for calibrating regularization, so we feel that our choices here are defensible. Nonetheless, we will modify the text to acknowledge that alternatives to both of these choices exist.

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")

That is true, but the drawback of regularizing an inversion by stopping the iterative algorithm early is that you are then dependent on the specific form of your iterative method for your regularization. This contrasts to explicit regularization, where the modeler can directly write down an equation for their regularization term and say, "I am penalizing this thing". We will rephrase this part to emphasize the difference between implicit and explicit regularization.

Methods

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

Yes, we will change this in the revision.

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.

Note that we did not shift the entire shelf column temperature by 10 C, only the surface temperature. The shelf base remained fixed to the melting point, so the shift in the average shelf temperature was less than 10 C.

You are quite right that the full inverse problem (for both rheology and drag) is underdetermined, as there are two unknowns at every point in the ice sheet. We wanted to make the problem more tractable by fixing rheology and only solving for drag. As we alluded to in the introduction and discussion, some authors have used inverse methods to solve for ice column rheology as well. However, the paper was quite long already, so we wanted to limit ourselves to only solving for a single variable.

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.

This method is our own invention. Reviewer 1 also requested elaboration on this point, especially regarding the behavior of the reordered mask under averaging and interpolation operations. We can add a short appendix explaining how the method works.

As far as terminology goes, we do not think it is strange to refer to velocity as a grid; we use "velocity grid" as shorthand for "a velocity dataset which is available in gridded format".

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

We will introduce Jr,c and Jd,c in the previous paragraph and reword this section in order to avoid confusion between smoothing in physical space and smoothing of the L-curve.

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.

That is true, the value of ½ of maximum curvature that we have chosen to bound the optimal lambda is arbitrary. Reviewer 1 referred to this choice as a "heuristic". In practice this will probably have a bigger influence on our ability to constrain lambda_max than on lambda_min. The drop-off in curvature below lambda_best is quite steep (Fig 5d), so alternate values of this heuristic threshold (say, 0.25 x max or 0.75 x max) will produce only small changes in the location of lambda_min. On the high side, however, alternate values of this heuristic threshold could potentially produce bigger changes in lambda_max.

We will edit the text to acknowledge that this threshold is an arbitrary choice, and to discuss the uncertainty that this introduces, especially in lambda_max.

Results

L365 : A very interesting point.

Thank you. Reviewer 1 suggested that we remove this point. However, we think that it is important to provide some context as to why previous authors may have found the results they did, especially if those results (an inability to find a corner of their L-curve in log-log space) conflict with ours. We are therefore going to leave this point in place.

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

We will remove this detail.

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.

It is true that model node count is secondary in the (Vogel, 1996) definition, however, in practical terms, the requirement that the regularized solution tends to the exact solution is also a requirement that the number of model nodes tends to infinity, since mesh resolution must approach zero in order to resolve the exact (continuous) solution. Furthermore, the requirement that the number of observations approaches infinity is also a requirement that the spatial resolution of the observations approaches zero, since mechanical glaciological inversions represent a single snapshot in time. Thus, the only way to increase the number of data points is for the velocity observations to have finer resolution. The Vogel definition of convergence applied to glaciological inversions can be summarized as, "in the limit that observational error and/or observational resolution approach zero, and in the limit that model resolution also approaches zero, then the inverted solution tends to the true solution".

Note that this definition of convergence or nonconvergence is not dependent on the forward models being a complete representation of the physics. Even Full Stokes inversions can be nonconvergent in the Vogel sense if the bed-to-surface transfer function (the "operator" in Vogel's terms) is sufficiently strong at attenuating short wavelengths. Conversely, SSA inversions could be Vogel-convergent if their transfer function only weakly attenuates short wavelengths. The key factor for this type of convergence is not the accuracy of the physics of the forward model; indeed, the Vogel analysis takes it for granted that the forward model is a complete representation of the physics. The key factor for this type of analysis is the roughness of the true solution in comparison to the attenuating effect of the operator.

We will rewrite and clarify this paragraph. In addition, this paragraph represents discussion more than results, so we will move it from Section 4.4 to Section 5.0.

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.

Indeed.

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

We computed best-fit exponentials to the graphs of lambda value vs mesh size. An exponential function can be written either as y=exp(kx), or y=exp(x/x0), with the e-folding length x0 related to the growth/decay parameter k by x0=1/k. In this case, the independent variable "x" is mesh

resolution, so our x0 is an "e-folding mesh size". We will rephrase this section in the revised manuscript to improve clarity.

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

We use the phrase "approximate bounds" to indicate the fact that the misfit metrics shown in Figure 9 worsen gradually as lambda is reduced between lambda_best and lambda_min, rather than having a single sharp threshold beyond which the misfit suddenly increases.

The point of this paragraph is that an inversion using a coarse mesh, evaluated at its own lambda_best, will produce structure that is similar to the structure produced by an inversion using a fine mesh at the same lambda value (even though this is not the lambda_best of the fine mesh). Thus, we can regard the inverted structure in the coarse mesh as having converged with respect to mesh resolution. The misfit increases as lambda is reduced from lambda_best to lambda_min, and then continues increasing below that. Thus, we can conclude that the space between lambda_best and lambda_min represents the approximate boundary between inversion results that have converged and results that have not converged. This analysis is separate from the convergence of the L-curve analysis, as the L-curve analysis involves changing lambda_best as a function of mesh resolution and lambda is held constant in this comparison.

We admit that this paragraph is awkwardly written and difficult to understand. We will rewrite it in the revision.

L555- (nonlinear sliding section). This is for me the most interesting result. What happens if you take your m=1 coefficient fields (C_1 (λ)) use the formula $C_3 |u|_3 = C_1 |u|$ to work out the corresponding fields for C_3 (λ) and then compute J_{reg} (λ)? 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)

This is a very interesting question. During preparation of the initial manuscript we performed these tests with our mid-resolution meshes (4 and 5) but we chose not to include them. We will run this test on our highest resolution mesh (1) and add this analysis in a supplemental results section.

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

Thank you.

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.

That is true, these two sections are our speculations about future work, rather than a discussion of the results we presented in this paper. As the manuscript is already quite long, we can remove these sections in the revised version.

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.

Our 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.

Brondex, J., Gillet-Chaulet, F., and Gagliardini, O.: Sensitivity of centennial mass loss projections of the Amundsen basin to the friction law, The Cryosphere, 13, 177–195, https://doi.org/10.5194/tc-13-177-2019, 2019.

Kazmierczak, E., Sun, S., Coulon, V., and Pattyn, F.: Subglacial hydrology modulates basal sliding response of the Antarctic ice sheet to climate forcing, The Cryosphere, 16, 4537–4552, https://doi.org/10.5194/tc-16-4537-2022, 2022.

Seroussi, H., Nowicki, S., Payne, A. J., Goelzer, H., Lipscomb, W. H., Abe-Ouchi, A., Agosta, C., Albrecht, T., Asay-Davis, X., Barthel, A., Calov, R., Cullather, R., Dumas, C., Galton-Fenzi, B. K., Gladstone, R., Golledge, N. R., Gregory, J. M., Greve, R., Hattermann, T., Hoffman, M. J., Humbert, A., Huybrechts, P., Jourdain, N. C., Kleiner, T., Larour, E., Leguy, G. R., Lowry, D. P., Little, C. M., Morlighem, M., Pattyn, F., Pelle, T., Price, S. F., Quiquet, A., Reese, R., Schlegel, N.-J., Shepherd, A., Simon, E., Smith, R. S., Straneo, F., Sun, S., Trusel, L. D., Van Breedam, J., van de Wal, R. S. W., Winkelmann, R., Zhao, C., Zhang, T., and Zwinger, T.: ISMIP6 Antarctica: a multi-model ensemble of the Antarctic ice sheet evolution over the 21st century, The Cryosphere, 14, 3033–3070, https://doi.org/10.5194/tc-14-3033-2020, 2020.

Vogel, C. R.: Non-convergence of the L-curve regularization parameter selection method, Inverse Probl., 12, 535, https://doi.org/10.1088/0266-5611/12/4/013, 1996.