To The Editor:

Please find below our responses to the reviews of Drs Morlighem and Cornford, with their comments in bold. We very much hope that these responses and the amendment to the manuscript satisfy yours and their concerns.

Thank you

D Goldberg on behalf of the authors

RESPONSE TO M MORLIGHEM

First of all we would like to thank the reviewer for such a thorough review, and for such supportive comments. We have attempted to address the reviewer's main issues below, by running additional experiments where necessary. As this response contains additional figures, we would like to state to the reviewer and the editor that, while the figures presented might be relevant to the response and certainly accessible to the editor and reviewer, we do not wish to present them in the actual publication as we believe the distinction between the experiment performed for this response, and those presented in the paper, is too subtle for most readers.

The comparison with the snapshot initialization is useful but a little bit biased in my opinion because it is based on an old (and probably bad) surface DEM (2002), with velocities that are 10 years younger. This exercise highlights the difficulty of using snapshot calibration because it is not always possible to find datasets that cover the same time period. But I would be curious to see the 30 year run with a ∼2010 snapshot calibration (with both surface DEM and velocity) rather than a mix of 2002–2010 data. It would not be surprising if this snapshot calibration, with better and more consistent datasets, yields results that are in a better agreement with the transient calibration. I understand that the authors want to compare the model output with existing data but it is not 100% clear whether the difference between the transient and snapshot calibration for the 30 year run is due to an inconsistency between datasets in the snapshot calibration or if it is just due to the difference between the two types of calibrations.

We attempt to address this issue at the opening of the discussion section by stating that the snapshot calibration is likely not the best possible given data currently available. However, there is a subtle underlying issue here: the importance of a model's ability to make a hindcast in order to make a forecast. These ice streams are in a transient state and have been for at least the last decade – and it seems logical one must capture this state in a model in order to address what the ice stream will do next. The method presented attempts to find this state by matching observations over some finite period. I don't believe there is any formal theory to say how long this period should be, but modelling investigations seem to point toward the decadal time scale (e.g. Payne et al 2005, Favier et al 2014). This then requires the use of less recent observations. Perhaps we are not saying anything with which the reviewer is not already in agreement. But it then must be shown that the same cannot be done with a snapshot calibration in the same manner, and this is the purpose of our snapshot calibration. We are careful to use the 2002 DEM, which is data, and not the 2001 DEM, which is backward extrapolation of data.

At the end of this letter, please find the results of a simulation similar to what you suggest (Fig 1 of this letter). We use the 2008 surface DEM from our transient data set, as this is considered to be the time point of the MEaSUREs data used by Seroussi et al 2014 for Pine Island. Note that the grounding line, initially retreated, proceeds to advance and there is a thickening signal close to the grounding line. We point out that the corresponding figure for the 2002-surface snapshot calibration (which we include in the letter to Dr Cornford) shows thinning near the grounding line, but not seaward-concentrated thinning. We also point out that in the 2002-initiated snapshot calibration, some ungrounding does begin to occur, and presumably would continue.

We would claim that dual questions of whether the difference between the 30 year runs is due to an inconsistency between datasets or if it is just due to the difference between the two types of calibration are in essence the same question, as it is very unlikely that data sets of diverse observations gathered even with the same timestamp would be perfectly consistent. The lack of strong thinning in both snapshot-calibrated runs is likely a transient pattern caused by the “shock” of initialising the model without fitting to a dynamic trend (though it is likely any nonphysical transients will not continue for more than a few decades).

This is more a comment than a suggestion, but I was somewhat disappointed that the model does not include a floating ice shelf downstream of the 1996 grounding line, because this is most likely the region where important processes (such as melting at the ice/ocean interface) triggered the acceleration and thinning that this region is undergoing. Ignoring this region and using boundary stresses as a control felt like putting all these critical processes under the rug. I understand the author’s rationale, but I would have loved to see the melt rates as a control and see if the transient calibrated model could tell us more about how the pattern of melting might have changed over the past decade (even with big error bars).

I realize it may feel unsatisfactory, but one must consider (as we have) what would be involved if ice shelves were to be included in the transient calibration. We include a more detailed consideration of this issue in the Discussion section.

At the very least, any trend inferred in the boundary stress parameters could indicate a change of the nature you mentioned (i.e. shift in buttressing due to ice shelf melting), though any inferences that can be drawn directly are qualitative. Inferring such a trend was an initial goal of this study; and given your comment we have elected to show it (see Fig 7), giving the caveat that we are not completely confident that J_{trans} has been lowered sufficiently to make this “real”. What is sought is a measure of the nature of an “adusted R^2” [e.g. Glantz & Slinker (2001), Primer of Applied Regression and Analysis of Variance, McGraw-Hill,New York], which can be used to determine whether one is overfitting or not when increasing the order of a polynomial fit. To our knowledge there is no such comparable measure in this instance. We are comfortable with this shortcoming, though, as the rest of the content of the paper does not hinge on this.

The other problem with having the stress at the grounding line as a control is that there is no unique solution (as mentioned by the authors p.4470), and it is even worse for the snapshot calibration. An increase in basal friction β has the same effect on the cost function Jsnap as an increase in normal stress σ. If, for some reason, the algorithm ends up with a σ that is too small, the model will artificially increase β in this region, generating an increase in basal friction right next to the grounding line (see Fig. 4b). With such a high increase of basal friction near the grounding line, it is not surprising that the grounding line does not retreat. Again, I agree that the transient calibration probably does a better job, because it is constrained by more datasets, but including σ in the control space will make the snapshot calibration worse than if the ice shelf was included.

We point out that were ice shelves included, this would indeed negate the need for boundary stress parameters but would then introduce melt rates, initial thickness, and Bbar as parameters, as mentioned above.

In a snapshot estimation, only Bbar would be required (as well as ice shelf thickness, which would need to based on BEDMAP2, and the issues with this are stated above). There is undoubtedly some “compensation” of the type mentioned above between Bbar and \beta^2, leading to some degree of equifinality which could lead to multiple solutions that lead to differing transient behaviors.

I concede it may be the case that the compensation between \beta^2 and boundary stresses is worse than compensation between Bbar and \beta^2 in a snapshot calibration. As the reviewer states, this may be less so in a transient calibration such as ours. But I believe that any such statements at this time are speculative; issues of equifinality cannot truly be addressed in this context without improved methods that allow Hessian characterization – a definite goal of future study as mentioned in Section 6.1.

Finally, I found the paragraph about the Rignot et al. [2014] paper not very convincing (but I might be a bit biased). First, I totally agree with the authors that Rignot et al. [2014] is based on a qualitative assessment and actual modeling is required to test this hypothesis. I also agree that when the fjords are narrow, the walls of the valley can exert enough resistance to prevent grounding line retreat along retrograde slope. Now, I am pretty sure that if melt rates were applied at the grounding line and in its vicinity, which is not the case here, grounding line migration would have been more dramatic. This is a very conservative simulation and provides a lower bound to the contribution of this region to sea level and grounding line retreat, and we cannot rule out more dramatic scenarios.

We point out that no time scale is formally given in the Rignot et al 2014 paper, and we do make the statement that rapid retreat past the 30-year horizon cannot be discounted. But it is a very good point that we do not force newly ungrounded ice with sub-ice shelf melting, and we now include this caveat in the section under discussion. (Though I point out there is no reason to assume they will be as high as e.g. PIIS, as this would depend on not only the newly created cavity geometry but that of the existing cavity which is largely unknown.)

I am not a big fan of the title for two reasons. First the term "near-future" is a bit vague, and "inferred" generally refers to the results of inverse modeling. This is just a suggestions and I will leave it to the authors to decide if they want to change the title but I would just take out the second part of the title: "Committed retreat of Smith, Pope, and Kohler Glaciers over the next 30 years".

We agree "near-future" is vague and have modified accordingly but do not see a problem with invoking a term from inverse modelling, as many studies featuring snapshot calibrations present their studies as inverse modelling as well. Macayeal (1992) used the term "inferred" in its title, and that study used a control method just as we have. "Transient Calibration" may in fact be a term new to ice modelling though "model calibration" is very well established in the computational physics literature, see e.g., Oden, T., Moser, R., & Ghattas, O. (2010), "Computer predictions with quantified uncertainty", Part I. *SIAM News*, *43*(9), 1–3.

p.4460 l.1: keep present tense "is calibrated".
Done, thank you.

• p.4460 l.12: I don't really like the term "steady-state" because snapshot inversions do not assume steady state (i.e. they do not assume that time derivatives are 0).

Yes, sorry, this slipped in here by accident. As the term is immediately explained we have simply removed this sentence fragment.

• p.4461 l.6: As such, (comma missing)
Done, thank you.

• p.4461 l.15: "ice thickness" is not really a surface properties. How about surface height?
Done, thank you.

• p.4461 l.18: "stiffness" generally refers to elasticity, viscosity might be more appropriate
Done, thank you.

• p.4462 l.4: you might want to cite Seroussi et al. [2011]
Thanks for pointing this one out.

• p.4462 l.19: integrated → run
Done.

• p.4463 eq.2: It is not really standard to add a factor of 2 for the constraints.
Apologies, this was a typo.

• p.44634 l.1: Minimizing J is not equivalent to minimizing J 0 , because otherwise you see that by taking Li > 0 and µi → −∞, we would achieve J 0 → −∞. We actually want to find the saddle point of J 0 .
You are correct. The wording is changed.

p.4464 eq.3: since basal friction opposes motion, you probably want a minus sign
this depends on how basal stress is defined, and the momentum balance is not stated in this paper; however, the formulation is consistent with Goldberg 2011, which is cited elsewhere.

• p.4464 l.16: the Lagrange multipliers are not the gradient of the cost function J generally. But they can be used together with the state variables to compute the gradient of J pretty easily.
"are then used to calculate"

• p.4471 eq.6: I am not sure to understand the equation, I would have defined the ice height above floatation as follows: hAF = s − R + min ρw ρi R, 0 (1)
Unless I am mistaken this expression can be negative where ice is floating, which the b_obs expression is required in order to determine, by saying that if ice is floating its hydrostatic base will be above R. But you are right that the additional expression for HAF was not actually given, it is now.

• p.4472 l.8: Thus, (missing comma)
Done.

• p.4472 l.10: I kind of disagree with this statement (if I understood correctly). When the grounding line retreats, the velocity increases over the entire domain instantaneously (e.g. Seroussi et al. [2014]). The only way to make sure that the inflow boundary does not affect the model is to go all the way to the divide, assuming that the position of the divide does not change. Now, given the time scale involved in this paper, I don't think the imposed flux affects the model significantly.
This is a subtle point. Just as in Seroussi et al 2014, we are not using a SIA model, and so of course stress perturbations are felt throughout the model instantaneously. But one would also expect stress transmission to be far less effective over strong bed. The region over which speedup is seen in Fig 5 of Seroussi et al is still quite fast-flowing, whereas velocities at the rightmost part of the domain are down to 100m/a or less, and the only parts of the domain that have low basal stress are the ice streams, as seen from Fig 4(b) of this paper. Thus I would not expect significant velocity change there (were I not imposing it) in response to frontal perturbation and GL retreat. Such a simple model as the kinematic wave theory I apply for this estimate is by no means accurate, just as it is not completely accurate when applied in other studies looking at the region. It is intended only as a back-of-the-envelope estimate as to how long thickness changes should take to propagate across the domain. Payne et al 2005 applied diffusive theory to the trunk of Pine Island Glacier to estimate the timescale of upstream propagation of thinning, which we now cite.

• **p.4472 l.15: Finally**
Copernicus's fault, not mine

• **p.4473 l.20: Thus, (missing comma)**
Done.

• **p.4474 eq.8: you probably forgot a factor 1/5**
I did.

• **p.4475 l.22: "rigour" → rigor (most of the paper is in American English)**
done

• **p.4482 l.3: Eq. B2 (no parentheses)**
done

• **p.4482 l.14: Thus, (missing comma)**
done

• **p.4483 l.14: Thus, (missing comma)**
done

• **p.4483 l.26: parentheses missing for references**
Done

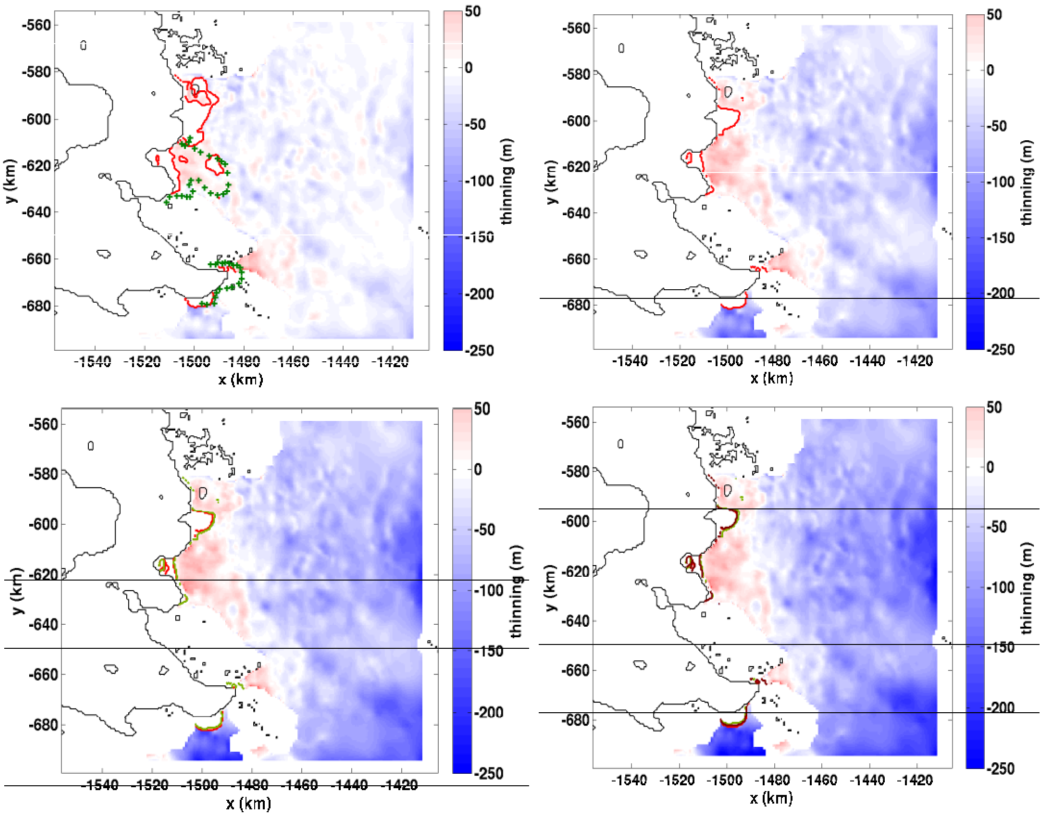

FIGURE 1: the equivalent of Fig 6 of the manuscript for an experiment initialized in 2008 with a snapshot calibration. The snapshot calibration uses MEaSUREs velocities, as in the main text, but with a surface elevation closest to 2008. Shading is thinning relative to 2001 as in Fig 6.Years of results are 2011, 2021, 2031, and 2041 corresponding to top left, top right, bottom left, bottom right. Ungrounding is widespread in 2011 but this is because it is imposed in the initial condition; the solution tends to ground downstream over time even as there is thinning upstream.

RESPONSE TO S CORNFORD

First of all we would like to thank the referee for a very encouraging and thoughtful review. Many good points are addressed and we feel this discussion will help to improve the paper. Note that there is a figure at the end of this letter in response to one of the comments below. We feel that its inclusion in the main text is not necessary as its "bulk" behaviour is given in Fig 5 of the paper and other aspects are explained in the text, and it might add "clutter" with 4 additional images -- but this is at the discretion of the Editor.

"It is possible that our snapshot calibration is equifinal" : that seems likely – if you have the same number of beta values as velocity values, and a normal stress too, then even in 1D there is a null-space. I think it is a vector made up from a perturbation to beta-squared one cell upstream from the face and a perturbation to the normal stress. But even if that were eliminated, there are a number of vectors associated with small singular values – such as oscillations in beta-squared some way upstream – which might end up being determined by the initial guess/choice of iterative method/regularization rather than data

I agree completely with the above statements. In this passage, however, I was referring to the possibility of compensation between heterogeneous control variables.

Still, the issue you bring up is a real one, and is wrapped up in the issue of subjectivity of priors – this is highlighted by Figs 4(b,c), which you mention below. As is commonly done, a smoothing term is applied to \beta in the cost function, the weighting of which for the snapshot calibration was chosen on the basis of bounding large variation over short length scales, and was applied to both snapshot and transient calibrations, there being no rationale to use different values. It is possible, however, that this prior information manifested differently in the different calibrations, leading to the "rib" features being more prevalent in one than the other. This highlights the subjectivity of such weightings and the need for minimizing subjectivity in prior specification, though we do not attempt to address this issue here. Mention of the above has been made now in section 5.2.

The additional information provided by the transient observations is sufficient to generate a better ice-stream state estimate" is a big claim in that case (not saying it is not true), but how does it come about? It seems to me that the transient calibration might work out better just because it matches velocity and surface in time. Put another way, the snapshot might be weaker largely because it mismatches, so that it insists on acceleration extending further upstream from the grounding line than it ought, which would look like a lighter pull (more buttressing) on a weaker bed.

This last point is a very good one, and one that had not occurred to us. This sentence from the text is now expounded on a bit further.

Stronger bed: Not uniformly stronger, though ? There is also an interesting ribbed structure in fig 4c (with a rib of strong bed close to where the GL seems to slow in the prediction).

Yes, as mentioned above (and now in the paper) this rib structure does arise, and it is difficult to know whether this is a result of different observations/different equations, or effectively less-strong smoothing. We do not make any speculation as to the physical underpinnings of the rib structures.

Negative buttressing: I like the idea that an Hσ that is larger than the non-ice shelf value might imply that the DEM h is too low. Might there be another explanation, too? That some parts of the grounding line are being pulled by faster flowing parts via the ice shelf. In that case you might expect the negative buttressing to line up with shear margins, which looks like it might be the case in fig 4a

This is a very good point and one which was considered but ultimately left out of the paper in favour of the other hypothesis. We now make mention of it.

Abstract: 'inverse methods'. This seems a bit slang to me.

We agree that this might be an overused or misleading term – particularly in instances where it has not been established that there is a unique minimum of the cost function in the control space. We have changed the mention in the abstract to the more descriptive "control method".

P4465, line 14-: The text doesn't actually say which method (AD, correct?) is used to compute the gradient of Jtrans. Is there space for a one or two sentence summary of the particular AD method?

This is explained in the reference (Goldberg and Heimbach, The Cryosphere, 2013) but we now say which AD tool is used.

P4467, line 27 : not so much the thickness, but the vertically integrated effective viscosity including crevassing etc.

Well, there is a distinction here between "estimate" and "control variable", I think you are saying I should limit this to the latter. But one could argue that both the control variables and the derived model state can be considered "estimates" which is why the sentence is worded as it is. In any event, the issue being addressed by this passage is now discussed in more detail in the Discussion.

p4470, line 1 ; 'high accuracy': maybe give numbers

done.

p4470, line 11, 'very weak bed' : perhaps give a number

done.

P4474 : line 8 : 'decreasing beta anywhere increases ice loss, lowering the bed only increases ice loss upstream of the projected 2041 grounding line.' is that quite correct? For the most part, there seems to be no sensitivity to beta downstream of the 2014 GL. The region where it seems to matter most and the bed does not looks to correspond to a grounded promontory in 2014. Is that bit lightly grounded?

If I understand the feature to which you are referring, it is lightly grounded, you can see this in Fig 6.

As for the region downstream of the 2041 (not 2014?) grounding line, we were bringing attention to the fact they are positive, albeit not as large as the negative sensitivities upstream. Certainly the sensitivities are small here – but a point made in section 6.1 is that on the whole bed elevation influence is marginal, or at least not overwhelmingly large.

Fig 6 : I'd like to see the same figure for the snapshot calibration. I'm guessing it has more even thinning?

Please see the figure on the following page. We have elected not to include this figure in the paper as we feel its point is made by Fig 5 – there is very little grounding line retreat. Your intuition is correct – thinning is less skewed toward the grounding line than in the transiently calibrated simulation, and is generally even aside from strong thinning in Kohler trough and, toward the end, strong thinning at the upstream boundary. This latter feature is likely seen because the input fluxes that were inferred from the transient calibration (and, we believe, minimize anomalous thinning) were not used in this simulation, and those that **were** used were too small.

We point out, though, that thinning and ungrounding does eventually occur downstream, and the slight thickening signal apparent at first does reverse. This could indicate that the anomalous thickening (relative and absolute) may indeed be a transient resulting from the snapshot initialisation, and the long term tendency is, indeed, retreat.

Fig 7: the legend has 'linear friction parameter', which I confused with a linear sliding law until I read the text properly. Maybe 'time dependent friction parameter'

Agreed, thanks for pointing this out.

Fig 9 (a): Do the upper schematics (the views of the front/join) add much? The planview could be larget without them

If you think this improves the understandability we are happy to make this modification, thank you.

Figure 1. The equivalent of Fig 6, corresponding to the snapshot calibration (calibrated to MEaSUREs velocities with 2002 surface DEM).