Review of:
E. Larour, J. Utke, B. Csatho, A. Schenk, H. Seroussi, M. Morlighem,
E. Rignot, N. Schlegel, and A. Khazendar:
Inferred basal friction and surface mass balance of North-East Greenland Ice Stream using data assimilation of ICESat-1 surface altimetry and ISSM

This manuscript presents the optimal estimation of a time-evolving state of the Northeast Greenland Ice Stream (NEGIS) using ICESat-1 altimetry and a transient version of the Ice Sheet System Model (ISSM). The optimal state is obtained through varying in space and time surface mass balance and basal friction (independent or control variables) in such a way as to minimize simulated vs. observed time-resolved ice surface elevations (dependent variable or objective function). Central ingredient to the gradient-based minimization is the adjoint model of the transient ice flow model, obtained via the operator overloading variant of algorithmic differentiation (AD).

The work presented is in my view an important milestone on the way to comprehensive time-evolving ice sheet state and parameter estimation. It demonstrates the feasibility of using time-resolved observations to constrain ice sheet simulations, and to systematically adjust uncertain, time-dependent forcings (if need be) to achieve a best-fit solution between model and observations. I have a number of comments, which I feel the authors can and should address before the manuscript is ready for publication.

Main comments:

1.
The first general comment is with regard to the choice of control variables relative to the model solved. Eqn. (1), (2) suggest that at every time step, a steady state momentum balance is being solved. The time-dependence enters exclusively through the continuity equation, expressed here as mass/volume conservation equation. This is common practice in ice sheet modeling, but the implication for formulating the control problem should be exposed:

Allowing for a time-varying alpha amounts to adding a time-varying source term in eqns (1), (2), but which are assumed to be steady-state equations. The authors should discuss the interpretation or implications of their approach. It seems to me that the model may be problematic in representing the impact of a time-varying alpha on a time-varying stress balance. This may explain why the optimization of J using the gradient w.r.t. alpha is of limited success. I don't expect the authors to make changes to their simulations, but to address this issue in the model formulation and in the discussion.

Related to this, I assume that the rationale for making alpha time-varying is that it might be physically connected to time-varying basal lubrication, e.g. through basal melt water (either via seasonal surface melt or geothermal flux or shear heating). I suspect that the main source of time-variability is the expectation that seasonal melt water at the bed would lead to intermittent (early-season) decrease in friction. This is supported by the discussion on p. 2353 (l. 2-14) of the relationship between basal hydrology and basal stress. However, this is not borne out by the inversion (see Fig. 6a,d). The question then is, what is the physical explanation for time-varying alpha? Alternatively, is the steady-state stress balance appropriate when using time-varying alpha?

Still related, I agree with the interpretation on p. 2349 (l. 14-19) of a

"clear equivalence between SMB and surface thickening rate, while basal friction is a direct forcing to the stress-balance equations (Eq. 1) and (Eq. 2), which have no direct bearing on the surface thickening rate", but think this statement needs to be stronger (to re-iterate):
The time-varying nature of alpha introduces a time-varying term in steady-state eqns (1), (2), a small inconsistency which the optimization my not be able to handle consistently.

2.
I caution the authors to refer to "improved" surface heights (e.g. caption to Fig. 6), or "improved" alpha, M_s. Fig. 6b suggests that S is not improved throughout. Whether all changes in alpha, M_s lead to "improved" values is not clear. A better term would be "adjusted", i.e. the optimal values of alpha, M_s are adjusted such as to yield a minimum least-squares misfit function J. In some cases the adjustment will indeed be improved estimates, in other cases, they will compensate for other model or estimation errors.

3.
p. 2342: The following statement:
"... showing a computation time for the gradient of the cost function with respect to either alpha or M_s on the order of 4 times the computation time for the forward model."
simply cannot be true, unless some very significant shortcuts have been taken. It is contrary to all accepted wisdom of algorithmic differentiation using operator overloading versus source-to-source transformation approaches for complex models. Please either revise this statement, or provide a description of which shortcuts have been taken, or provide a model setup that enables testing of this statement by outsiders. (Even if that factor should turn out to be much larger than 4 times, the author's achievement is still very significant).

4.
p. 2342: l. 17/18:
The sentence :
"the fact that we do not rely on the adjoint-state but rather on AD to compute the gradient, and that the inversion is temporal in nature."
is unnecessary and wrong (or a misconception of what AD does). The code generated via AD *does* compute the adjoint state at each time step (no matter which form of AD is used). Therefore, you *do* (have to) rely on computing the transient adjoint state. The only thing you have avoided is having to hand code the adjoint model of your time-varying model that computes this state. AD is only a shortcut for avoiding hand-coding the adjoint model, not a shortcut for avoiding computing the adjoint state.

5.
p. 2345: l. 18:
The sentence:
"Assimilating altimetry data into a forward transient ice flow model presupposes that the model itself is spun-up in a way that more or less closely matches observations for the time period considered."
is misleading or wrong. Nothing prevents an assimilation problem to be formulated in such a way that initial conditions and model parameters are adjusted such as to correct a poorly spun-up initial state (e.g., Goldberg and Heimbach, 2013). In fact, "data assimilation" in its most common usage in numerical weather prediction (NWP) is synonymous with finding initial states which lead to optimum fit to observations at analysis time (and optimum forecasts).
A more accurate statement might be:
"Since our assimilation method does not adjust initial conditions of the model, we have to rely on a spun-up model state which more or less closely

matches observations for the time period considered. In general, the success of inverse methods applied to nonlinear problems often relies, in practice, on initial guesses of the independent variables that yield states that are not too far from observations."

6.
p. 2347: l. 20/21:
"Because the model spin-up does not reach a configuration that matches the altimetry time series within a 1 standard deviation, we are still forced to adjust the overall mean of the entire altimetry time series so as to center it on the modeled surface height in 2006."
I'm not sure I understand what this means. I think what is being said is that a time-mean bias (spatially constant or spatially varying?) is removed such as to obtain a better initial misfit? This needs to be described more clearly so it is more transparent to readers what is being done. Ideally, a figure should be added, depicting the true mismatch without the adjustment.

7.
p. 2347: end of section 3.2:
A description is needed regarding the exact nature of the time-variation of alpha and M_s. Is the period between two consecutive adjustments the same as the model time step (i.e. two weeks), or is it longer-period? This has repercussion on the dimensionality of the control vector. if Nx*Ny is the dimension of a 2-D field, then the control space would have dimensionality Nx*Ny*nUpdates. nUpdates could be either the number of time steps (roughly [2009-2003+1]*365/14), or a coarser partition of the integration period.
Another question is why the inversion for alpha and M_s have been performed separately (l. 19,20). A formal inversion would invert for both parameters jointly.

8.
p. 2348: l. 2/3:
I am not sure how it can be inferred from Fig. 4 that
"best-fit to observations can only be improved by varying forcings over the entire space and time domain."
All that Fig. 4 shows is that the gradients are space-time dependent. This, in turn, is a consequence of the nature of the observations. To see this, note that for a cost function of form:
   J = 1/2 (F(x) - obs)^2,
the gradient is of the general form:
   dJ/dx = (dF/dx)^T * (F(x) - obs),
i.e. the gradient is "driven" by the (linear) model (F(x)) vs. data (obs) misfit. To the extent that (F(x) - obs) is time-space varying, so will be the gradient.

Related, p. 2348, l. 12-15 and p. 2349, l. 5-7:
"For dJ/dalpha, this can be largely explained by the fact that basal friction is much higher there than near the coastline, making it much harder for equivalent variations in basal friction to impact ice flow dynamics and surface heights."
This may be the case, but is an interpretation not readily borne out by the analysis. The simplest explanation that is supported by the analysis is the same as above, i.e. the fact that
   dJ/dx = dF/dx * (F(x) - obs)
implies that for small misfits (F(x) - obs), which is the case inland, the gradient is small, no matter what the size of x (here = alpha), unless (dF/dx)^ itself would be very large (but which too would require demonstration) .
Linking the smallness of dJ/dalpha to the largeness of alpha itself requires further scaling analysis.

Still related, the "controlling mechanism" invoked on p. 2349 (l. 5-7) can instead be simply explained by the small residual model-data misfit in the regions suggested upward of the suggested demarcation.

9.
p. 2350, l.9 onward:
Figures 6c, g suggest that the optimization "corrects" winter mass balances for both positions I, II to be solidly negative, compared to their first-guess values which are near zero or slightly positive. Is this expected? The implication would of a negative mass balance not just during summer months but throughout the year would seem significant.

Details:

* p. 2332

l. 20:
It seems more prudent to refer to the common terminoligy Global Mean Sea Level (GMSL) rise. Alternatively, refer to "sea level change", since regional sea level trends may be negative (i.e. sea level drop) over the last 20 years.

l. 21:
Update to IPCC AR5 (plus relevant reference)

* p. 2335:

l. 14-21:
correct all section numbers (section N -> section N+1)

l. 15:
Here and throughout the manuscript (e.g., p. 2336, l.5; etc.) it would seem "nicer" and consistent with the estimation/control theory literature, to refer to "objective function" instead of "diagnostic".

* p. 2336:

l. 8:
Replace "Ice flow on the NEGIS" -> "Flow of the NEGIS"
(seems to me that the ice doesn't flow *on* the NEGIS, and "Ice flow" of the "Ice Stream" seems redundant).

* p. 2339:

l. 2:
The cost function sums the SQUARED differences.

l. 16:
Here, and later in the manuscript (e.g., p. 2342) the notation
J = F(alpha(t),M_s(t)) is not well defined, or misleading.
If F indeed refers to the model (defined how? I guess the system of eqns. (1) to (6)) then J is not scalar-valued. Instead, I think what you mean is:
J = J(F(alpha(t),M_s(t)))

* p. 2340:

l. 8:
Replace "adjoint theory" by "adjoint method"

* p. 2342:

l. 8:
Reword:
"... we can AD-compute dJ/dalpha, gradient of..." to
"... we can compute dJ/dalpha, the gradient of..."

l. 12:
It might be more conceptually more transparent to distinguish between first-guess alpha_0 and optimized alpha = alpha_0 + Delta alpha, i.e. write:
"... we can infer an update Delta alpha to alpha_0, such that alpha = alpha_0 + Delta alpha leads to a simulated surface height evolution that minimizes the cost function".

l. 13/14:
Not the "inverted" alpha itself best fits the data, but the state computed with the adjusted alpha does.

l. 22, 24/25:
"Here, we do not assimilate both forcings alpha and M_s."
This statement is wrong, it mixes up dependent and independent variables. Observations are assimilated, not input variables. What you mean is either: "we do not invert for both ..." or "we do not adjust both ...".
   Likewise, the sentence:
"which parameter assimilates existing altimetry observations most efficiently"
is ill-worded.

* p. 2343:

l. 6:
"by a simultaneous reconstruction of the surface topography"
Use of *simultaneous* makes sense for reconstruction of A *and* B. A is surface topography. What is B? Otherwise drop *simultaneous.

* p. 2346

l. 1/2:
"instantanoeus spin-ups"
This seems a bit of an oxymoron (or the term "spin-ups" misleading), so perhaps add "or snapshot inversions". Also, in the following reference list, it seems warranted to add Petra et al. (2012).

l. 6-9:
It may be true that:
"However, this approach relies on a steady-state thermal regime for the ice sheet, which is not realistic, ..."
but the same is true for the approach presented here, see p. 2888, l. 3-5:
"The thermal regime of the ice is not captured in our transient ice flow model ... We believe this approximation to be realistic".
It would seem that the approximation made holds equally well in both cases. By the same token, the statement "usually leads to lumping any mismatch between model and observations into the inversion itself" is equally valid in both cases, to the extent that it refers to the thermal regime.

l. 21:
"followed by a relaxation of the ice sheet/ice shelf over a period of 50,000 years"
I am not sure what this means, or whether this is a common numerical method.

I suggest describing what the "relaxation" involves (in fact, some authors refer to "relaxation" as a simple form of data assimilation, but I suspect this is not implied here?).

l. 22:
"The climate forcing is constrained by an SMB taken equal to ...".
I don't understand what is meant here by *constrained*. I suspect the authors simply mean: "The climate forcing is represented by the time-mean SMB between 1971 and 1988".
Similarly, it is somewhat unclear to me why the period 1971-1988 is chosen as "climatology". The Box et al. (2013) time series goes back to 1840, so why not taking 1850-1988 as a more representative climatology (i.e. a better average over decadal variability), or any other start date between 1840 and 1971? If the 1971-1988 time-mean is used for the integration prior to 1971, it would seem more likely that SMB undergoes an artificial jump in 1840 (the time at which the Box et al. time series is applied) than using a time-mean SMB which is more representative to 1840(?)

* p. 2348, l. 28 / p. 2349, l.1/2:
I have difficulties seeing the "clear demarcation line" and "abrupt transition in ice thickness". I'd suggest adding corresponding isolines/ contours to Figs. 4 and 7 that delineate the transitions in question.

* p. 2349, l. 19/20:
Reword "... between both methods ..." to
"... between varying alpha or M_s ..."

* p. 2353, l. 1:
"... exhibits high variability ..."
In space or time, or both?

* p. 2353, l. 28/29 and p. 2354, l. 3-6:
"Here, we propose..."
A good proposition, one that has already been formulated by Heimbach and Bugnion (probably others before), and that has already been explored by Goldberg and Heimbach (2013), who used time-varying altimetry and surface velocities with inhomogeneous temporal sampling (to reflect heterogeneous InSAR vs. ICESat sampling, albeit in a synthetic experiment) to constrain a transient ice flow model and simultaneously infer best-estimate initial conditions and basal sliding.

References:

N. Petra, H. Zhu, G. Stadler, T. J. R. Hughes and O. Ghattas, 2012. "An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model". Journal of Glaciology, Vol. 58, No. 211, pp. 889-903.