Interactive comment on "Towards a better ice sheet model initialisation and basal knowledge using data assimilation" by Cyrille Mosbeux et al.

Answer to *R. Arthern (Referee)*

I think this manuscript could be published in Geoscientific Model Development after a few changes. This paper provides a comparison between two different approaches for estimating the initial state and parameters of an ice sheet model. The paper provides an illustration of simultaneous inversion for bed slipperiness and bed elevation using adjoint methods. This part is not an especially novel endeavour in itself, but it is used
here to provide a reference for another method, the combined adjoint/nudging method. The merits of the latter approach lie more in its ease of implementation than its theoretical justification. Nevertheless, if adjoint/nudging is shown to be competitive with more complicated approaches, as seems to be the case here, this would represent a valuable service to those ice sheet modellers that presently have the wherewithal to
invert for basal drag coefficient, but have not yet considered the shape optimisation problem of recovering the basal topography.

The paper is well-structured and clearly written. The figures are useful and clear. The two parameter adjoint approach has perhaps been described better elsewhere, but I think the examination of the combined adjoint/nudging approach as described here is probably still worth publishing in GMD. The results are sufficient to support the interpretations and conclusions. The authors make clear which parts are new, and which have previously appeared in the literature. The title is OK, although the new feature of the paper is the combination of adjoint-based inversion and nudging and this is not prominent in the title. The abstract is fine. The mathematical presentation is clear enough. The number and quality of references are OK

We revised the title in order to give more details about the assimilation methods used in the paper. Therefore, the new title is : "Comparison of adjoint and nudging methods to initialise ice-sheet model basal conditions".

The simulations used to illustrate the comparison are undoubtedly highly simplified: a simplified approximation of the stress state is used, a 2D flowline rather than a 3D ice sheet is considered, and all the measurements considered in this manuscript are synthetic. These simplifications are expanded upon below. However, to my mind, these do not detract from the central purpose of the manuscript, provided that it is recognised that this paper provides a necessary test that should be passed by the adjoint/nudging method, rather than a sufficient test that would guarantee its usefulness by other models in more general circumstances. In short, this paper might motivate readers to consider the adjoint/nudging method for initialising their models, but each modeller will still need to demonstrate that the method works for their model, in 3D, not just 2D, and each modeller would preferably test the approach with real observations as well as idealised 'twin' experiments.

The simulations use the Shallow Shelf Approximation SSA. This is the shallow aspect ratio limit appropriate for flow over a very slippery substrate. For shearing flow over non-slippery sediment, another commonly used limiting approximation, the shallow ice approximation (SIA), which is not used here, would be more appropriate. Nowadays, the practical initialisation problem for ice sheets is more likely to be performed with a more sophisticated stress-balance using a vertically integrated 'hybrid' blend of SIA and SSA stress states, or a depth-resolved higher-order model, or Stokes flow. In these more

sophisticated models, the transition from slippery to non-slippery substrate poses no special complications, while for the SSA approximation used here, the accuracy of the model will deteriorate whenever the assumption of extreme slipperiness is violated.The paper would be improved significantly if similar twin experiments were performed using the adjoint/nudging approach for a hybrid model, a higher order model, or a Stokes flow model. This would be especially valuable if it turned out that the bed recovered from the inversion was shown to depend on the approximations used in the momentum equations. The chief selling point of the combined adjoint/nudging method is that it would be easy to apply to more complicated models, so I am not sure why this is not done in this paper. As it stands, the paper points to the promise of this approach for initialising more complicated models, but without a relevant example, it is hard to know whether this is real promise or false promise.

A relevant point, which arises in the upper paragraph, concerns the potential added value to the article if the adjoint/nudging approach was used with a higher order model or a Full-Stokes model. Indeed, the purpose of the adjoint/nudging approach is its easy adaptation to such models. However, we only considerate the SSA approximation since the point of the paper is to compare this approach to a more conventional one (the inversion of the two parameters using adjoint method).

Concerning the impact of the use of higher order model on the bedrock and basal drag solutions, it is clear that the solutions will depend on the direct model. However, this was not the main purpose of the paper although the idea is relevant.

Also, our case of reference is a steady-state case constructed using the SSA approximation. Using a Full-Stokes direct model (or others higher order models) would change the reference case since the reference friction or the reference ice surface velocities are adapted to SSA. These changes could make harder the comparison between the different inversions.

Moreover, the use of adjoint/nudging in full-Stokes model is one of our next goal as well as 3D applications where the results will be compared with the results using SSA.

I have fewer concerns about using a 2D flowline simulation for illustrating the two methods, but some readers will wonder whether the two methods would still perform comparably in 3D. The paper is still quiet compact, and a 3D example would make for a fuller investigation. The paper states that the methods can be applied in 3D, but it would be better to show an example. The use of exclusively synthetic observations represents a limit to the information provided by these simulations. To the authors credit, the data used are based on a real flowline, so the bed inversion at least can be checked. There are a number of regularisation parameters in the inversion(λ_{α} , λ_{zb} , T, k). Inevitably, these parameters represent rather vague prior information and are quite poorly constrained (Arthern R.J., J. Glaciol., 61 (229), 947-962, 2015, doi: 10.3189/2015JoG15J050). At least it would be good to include a table showing how much the inversion of the bed can vary from the 'true' bed when these are varied.

The purpose of the paper is to illustrate the performance of the different methods in a synthetic case. Although, the algorithms are implemented to work in 3D (pseudo-3D for ATP) we made the choice to keep a 2D example which appears more didactical from our point of view.

Concerning the regularisation parameters used in the inversion, the choice of these

For ten year simulations with the forward model, it should now be possible to test the evolution of the surface against real altimetric observations. This is perhaps too much to ask of an initial demonstration paper such as this, but time series of elevation data for Jakobshavn are available and it would be interesting to see how well the different methods reproduce the actual behaviour over ten years.

In reality, the twin experiment is inspired from the true observations such as the bedrock shape in order to make it more realistic. However, the case is synthetic. The reference friction is inferred using adjoint method, optimising the misfit between observed and modelled velocities, after what we have conducted a model relaxation up to steadystate. The reference velocities and the rate of change of ice thickness (equal to zero) correspond to this steady-state.

Moreover, the flowline behaviour can be very different that the 3D behaviour of a glacier which is more complex. Therefore, although the idea is very interesting, it would not be very relevant to make a direct comparison between our surface evolution during 10 years and the current evolution of the Jakobshavn Isbrae glacier.

In summary, this paper makes one point quite nicely – that the adjoint/nudging approach can work well for the SSA, for flowline models, as judged by synthetic 'twin' experiments, but it still leaves many avenues to be explored.

Thanks for this positive comment. It leaves many avenues to be explored such as adaptation of the method to a 3D real cases. As we already mentioned, this is our next goal and we really are looking forward to show it soon.

Minor points:

Line 46: Replace 'the' with 'then'

Changed in text.

Line 49: Replace 'constrain' with 'constraint' and 'are solution' with 'are a solution'

Changed in text.

Line 125: Are unweighted least squares cost functions such as these appropriate, or should error covariance weighting be applied? Might be worth some discussion.

Actually, we can add an error covariance weighting. In this « idealised » case we do not favour some dhdt_obs with respect to others since measured dhdt_obs are considered perfectly known. Of course, in real application we could add some covariance weighting by using some error estimations on observations such as we can found in Flament and Remy. 2012. However, these estimations seems hard to get.

Note that we tested the effect of noise on dhdt_obs. Firstly, by adding Gaussian white noise to dhdt_obs (see line 370) with no significant effect on the results of the inversion. Of course systematic bias have an impact and the introduction of an error covariance weight

in the cost functions could be helpful to reduce this impact (by reducing the weight of the points where bias is known to be high).

We specified in the new version of the paper that covariance error can be taken into account and why we do not use it in our case where we have perfectly known observations or vitiated with Gaussian noise (see line 135).

Line 175: Give more details of the Gaussian used to define *k*. How do results depend upon this choice?

Of course the choice of the variance of the Gaussian will impact the results. Tests show that excessive variance values induce unphysical call-back amplitudes when departing from observations. After a few cycles, the resulting bedrock induces an increase between modelled and observed velocities that the basal drag inversion is not able to overcome.

In our specific case, the threshold on the variance, between excessive and acceptable misfit on velocities is slightly above1 km.

Under this threshold, the value of the variance have little impact on the final result in term of cost functions. However, we tested a few values (0.2, 0.4, 0.6, 0.8 and 1km) among which the variance of 1km gives the best agreement between misfit on velocities and misfit on the rate of change of the ice thickness.
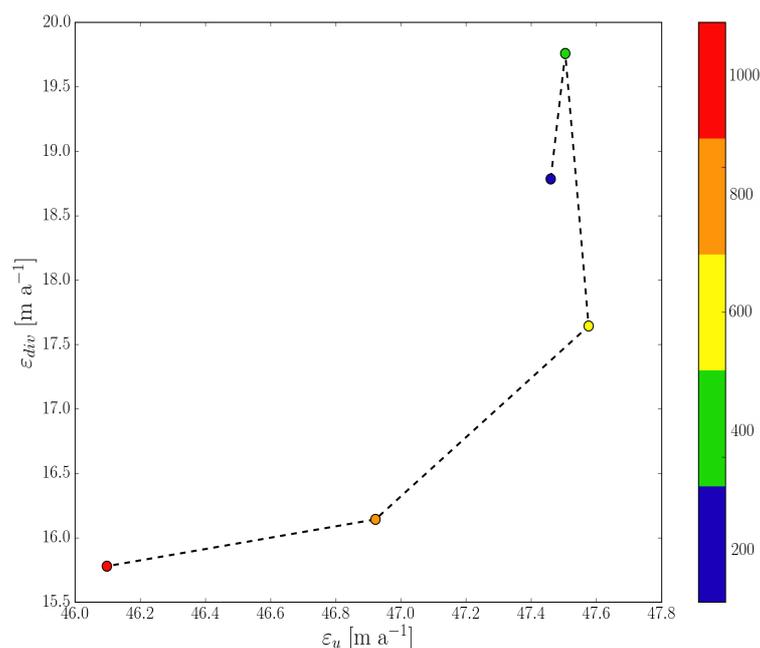
Fig. Misfit on ice surface velocities and on the rate of change of the ice thickness with respect to the variance of the parameter *k (200m, 400m, 600m, 800m, 1000m).*

Line 210: Jakobshavn is misspelt

Changed in text.

Line 267: Sometimes the 'L' of the 'L-curve' is very clear, sometimes not. It would be good to show the two cross sections through the 'L-surface' at the chosen values using a log-log scale.

In our case, the 'L' is visible by looking at the 3D-Lcurves (Jv, Jdiv, Jreg) and (Jv, Jdiv, Jzb). However, it is more an « optimal area » than a specific point. The 3D graphs are not so easy to read if we cannot navigate in the 3D space. It seems more easy to show the « specific point » we selected by plotting two scatter plots : (Jv,Jdiv) and a color for each dot indicating the value of Jreg and the value of Jzb.
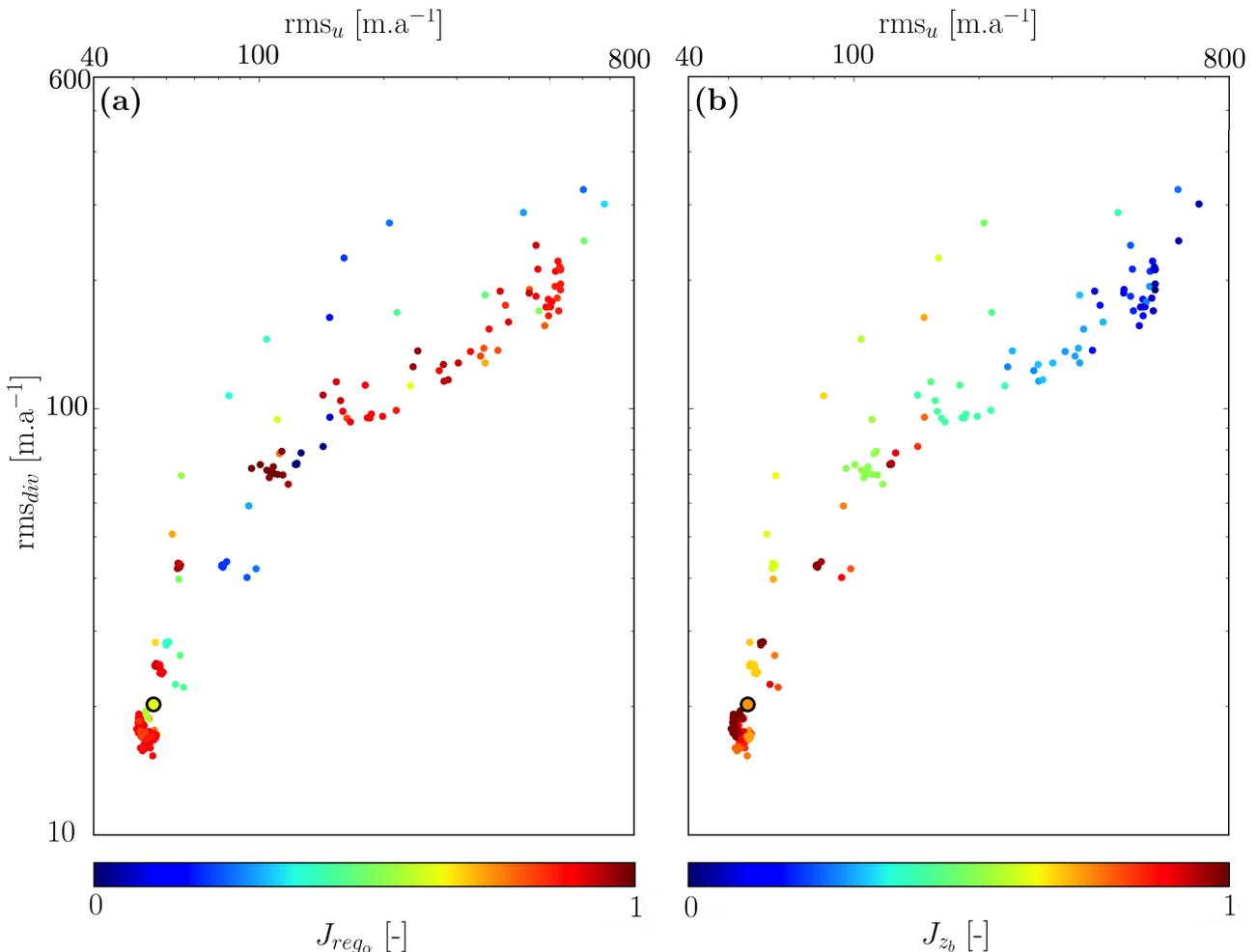
Fig. *Average velocity misfit on the horizontal axis, average misfit on the ice thickness rate of change of the divergence on the vertical axis for each of the 255 couples of regularisation. Values for (a) Jreg and (b) Jzb are represented on the color axis. The optimum is marked with a black circle.*

The graph shows that the dot marked with a black circle, corresponding to our optimum ($\lambda_\alpha = 1e11$, $\lambda_{zb} = 1e7$), is a good compromise between each cost function and allow to minimise both misfit on the divergence and ice velocities while keeping a good agreement with the *a priori* on zb and the regularisation on ß.

As suggested, the figure above has been added to the article to support the choice of the regularisation parameters. It also has been referenced and discussed in text.

Line 305: Since *T* has in effect become a regularisation parameter it would be good to comment whether this trial of a few values is consistent with the treatment of the other regularisation parameters – is there an equivalent to the 'L-curve' for choosing *T*.

Indeed, the choice of *T* can be done similarly to the choice of the regularisation

parameters since there is a relation between *T* and Jv and Jdiv. Nevertheless, the choice of *T* seems less straightforward in the present case. In the Figure below, we plotted T in function of the average misfits on surface velocities and on the thickness rate of change. We can also follow the scatter plot approach we have used for the ATP. Both graphs are given below.
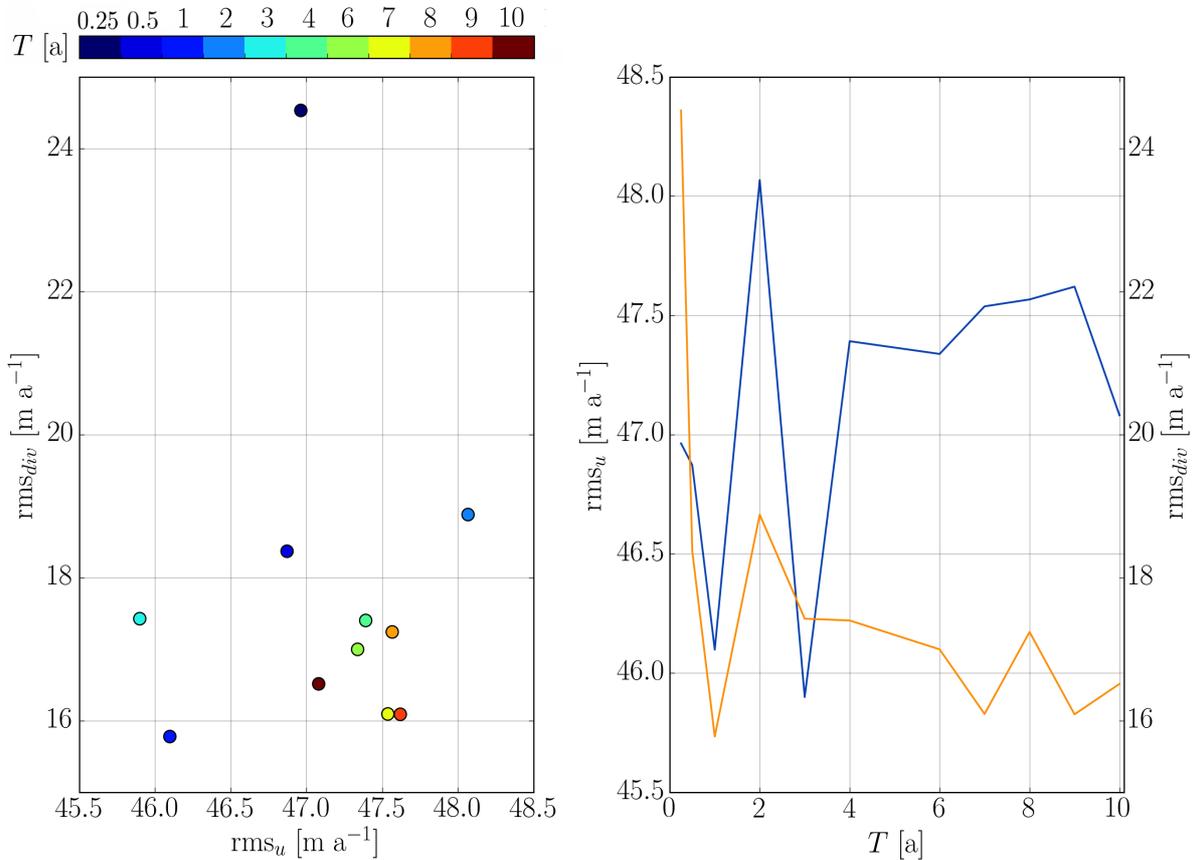
Fig. *Misfits on velocities and ice thickness rate of change with respect to T. (left) Scatter plot where the color correspond to the size of T. (right) Blue curve corresponds to the velocity misfit and orange curve correspond to the ice thickness rate of change misfit.*

Here, 10 ANC cycles are conducted. The figure shows that *T* periods over 4 years tend to largely increase the misfit on Jv since there is no control on velocities during nudging (see sec. 4.2). *T* periods under one year generate high misfit on Jdiv since *T* are not long enough to significantly reduce divergence (of course more ANC cycles would solve this last problem but at the expense of the computational cost). Note also that very long periods, such as 20 years seems, unable to converge since it induces too high ice thickness evolutions in the early cycles of ANC.

The scatter plot shows us that the 1 year T periods gives a good agreement between fitting velocities and divergence. However, the choice of periods of 1 (minimum for Jv) or 3 years (minimum for Jdiv) is still debatable but the choice of a shorter *T* period is, again, to the advantage of the computational cost, which can become a significant choice criterion in cases of larger domains (extension to 3D cases) or more refined meshes.

We decided to not add this Figure to the paper but to detail our choice and our method of selection for the *T* period in sec. 4.2.