Response to Referee #2: "Interactive comment on *A comparison of two Stokes ice sheet models applied to the Marine Ice Sheet Model Intercomparison Project for plan view models (MISMIP3d)*"

General Comments

This paper presents results for the MISMIP3d experiments for the recent finite-element Stokes flow model FELIX-S and compares them with revised results from the well known Elmer/Ice model, (that have already been published in TC by Gagliardini 2016 et al). It is good to see a second high resolution Stokes model performing these prognostic experiments, and it looks as though the two models agree to within the published truncation error of Elmer/Ice, and indeed, seem to have similar truncation error. Notably, both models [show] a steady state grounding line position about 80 km upstream of the SSA models, as seen for Elmer/Ice in the original MISMIP3d results and as the truncation error at an along-flow mesh spacing of 50m is around 5km, the paper adds weight to the view that this difference is due to the more complete physics in Stokes model than numerical error.

We thank the reviewer for his comments and for pointing out the model similarities in the context of the published Elmer/Ice truncation error. In the original version of our manuscript, we did not explicitly point this out. In our edited version, we have made it more explicit in several places that, when run at their highest grid resolutions the two models discussed here show agreement to within the published Elmer/Ice truncation error of 2-3 km.

Response to Major Comments

It seems that the difference between the models is entirely truncation error, ie both models appear to implement the same physics. At several places the paper suggest that the difference between the models tells us what the error in those models is. That's not quite right – it tells us what the larger truncation error of the two models is, but not which model has that error. Imagine that Elmer/Ice had a 5km error in the GL position and Felix-S had 100 m error (I know that's not true – this is just an illustration). Would the Felix-S developers be happy to accept that the error in their model was 5 km on that basis?

It is not enough to compare two Stokes models: we need to compare models that have known truncation error. On top of that, if we know the truncation error (and that is the major issue), then the comparison is much less important than this paper suggests in several places including the abstract and conclusion. So while it is certainly desirable for future MISMIP type exercises to include multiple Stokes models, the proposal that it be a requirement, is, I think, an opinion rather than a fact, and should not appear in the published paper.

We take the reviewers point, and have updated the relevant discussion in the abstract, the discussion, and conclusions sections of the paper. We now state explicitly that, at high resolution, the differences between the two models appear to largely fall within the previously published estimates for the truncation error of Elmer/Ice. While we agree that this diminishes the requirement that multiple Stokes models participate in future intercomparison exercises, we do include a strongly worded opinion that this should still be the case, simply as a way of confirming that in future studies.

There should be a convergence study of the Stnd experiment run by FELIX-S with respect to along flow mesh spacing dx. Other papers (e.g Seroussi 2015) make it clear that the Stnd experiment is harder to get right than the reversibility. Gagliardini 2016 included such a study and I think this paper should too. That is, run the Stnd experiment at dx=100m,dx=200m,dx=400m. etc in addition to dx=50m(dx= 25 might be too expensive). That should allow an estimate of the truncation error as a function of dx (it will presumably be about 100 dx, as it is for Elmer/Ice) – this is already hinted at by the 5km difference in the two dx = 50m experiments mentioned on page 8.

As suggested, we now report on the results of the grid convergence study for the Stnd experiment. There is a new paragraph discussing this at the end of section 5.2 and a new figure showing the results of the convergence experiment (new Figure 4).

Response to Minor Comments

Fig 7: The caption talks about an absolute value of x (ie |x|) but there are negative values on the plot. Also, what does the least-squares line tells us about convergence? I would expect |x| ~ dx = 1/N_y. I suggest plotting the |x| differences on a log-log scale.

We have replaced the previous version of Figure 7 with a newer one in the spirit of the reviewer's suggestion.

P8, L20; "reversibility" is not better just because the GL is closer to its initial point after 100 years. It takes ~10^4 years to see the return to steady state. What's important here is the convergence of the FELIX-S and Elmer/Ice solutions within themselves and with each other.

We have added a clarifying comment to this effect in the revised version of the manuscript.

P9, L25: likelihood has a specific meaning. Maybe use 'tendency' or something else?

In the revised version of the manuscript, the relevant sentence has been changed to: "This in turn *will make it more likely* that neighboring nodes may also eventually ground."

P11, L2, and elsewhere. Uncertainty? I would say error, numerical error, or some such. Uncertainty is quite a broad term suggesting probability , but here we know that

differences are due to (deterministic) error.

This part of the conclusions (formerly p.11) has been re-written. Elsewhere in the paper where we formerly used the word "uncertainty", we have changed it to "error", "model error", etc.

P11, L4: Future efforts. I don't really see the value of comparing identical treatments. I guess it might expose bugs, or differences due to solution method (as opposed to discretization methods), but it's not usual to publish that sort of thing (even if bugs are a common enough occurrence).

We have removed this particular section of the final paragraph. Now the "future efforts" discussion focuses on trying to identify if the different methods used by FELIX-S lead to different or similar truncation errors relative to those found for Elmer/Ice.