I have read the response and the new manuscript. I am not really sure what to recommend, because it seems that the authors do not agree with the broad scope of both reviewers earlier comments. To summarize, I do think that the revised MISMIP3d results should be published and I do agree that Elmer/Ice should be considered the reference ice sheet model until somebody produces something as good. But I do not think that the comparison of the LG, FF and DI methods is the best way to discuss Elmer/Ice numerical error (I will give reasons below). I think the correct method is a convergence study, which is included among the results, but on its own a convergence study of a well-known idealized problem without new methods seems a bit thin for a publication. I suggest that the results should be worked into a brief communication.

I will respond to some of the authors general response points (emphasised) below.

The fact is that Elmer/Ice results for MISMIP and MISMP3d are the only ones produced using a full-Stokes model, and as such, they are often used for comparison with lower order models (e.g. Feldmann et al., 2014) and as a reference during the development phase of new Stokes solvers. It was therefore important for us to let know the community that these results are sensitive to the way the friction is implemented in the close vicinity of the GL, at least for the resolutions that were used to produce these results. By publishing this paper, we also offer the community new Stokes MISMIP3d results at a much higher resolution than those previously published.

I agree that the new higher resolution MISMIP3d results and error estimates should be published, especially given the status of Elmer/Ice as the Stokes model with the strongest history of publications.

We agree that the initial title was overstating the results presented. We agree that it is a question of mesh sensitivity, but, we still think that the differences are substantial for the mesh resolutions that were used to produce the MISMIP3d results and they will certainly be substantial for the mesh resolutions that might reasonably be used for a real large-scale application.

Perhaps the isssue here what is meant by 'substantial'. Clearly, switching from (say) LG to DI, the steady grounding line position differs by many mesh spacings, which is perhaps what the authors mean by substantial, and is a reasonable definition. I had in mind that for resolution dependent estimates of a single number to be substantially different, the difference had to be clearly larger than the error estimates – and that can only happen if the error estimates are different.

We know for sure that finite element results are mesh sensitive, and we show in this paper that by increasing the mesh resolution, the difference between the three methods is decreased. But we also know that such small resolutions are often not tractable for 3d problems. By quantifying the differences obtained with the three different methods at different resolutions, we advise the community about error associated to FS results for a given resolution.

If the intention of the paper is to quantify Elmer/Ice numerical error in MISMIP3d, then I support that, but I maintain that the differences between LG, FF, and DI don't themselves tell us much about it. Imagine that only a single resolution had been run : in MISMIP3d there would be no information which had the largest error. The only error measure error you would have is the difference between retreated-to and advanced-to steady states in MISMIP, which is not broadly applicable (imagine doing such an experiment for PIG, given the computational time that was needed for a 50 year run). It is only by considering several mesh resolutions that we can see that DI is incrementally better than the others.

So what is the value of comparing LG,FF, and DI?. You could say that in future realistic runs you might use DI, and estimate the size of the numerical error by also running LG or FF. In other words, you have a method for examining numerical error. I don't think that would be practical. First, as noted with respect to the Tsai and Leguy rules, all three would in some cases give the same answer and say nothing about the error. Second, if the cost of running at mesh resolution Δx is 1 unit, we can be confident that running at $2\Delta x$ should be 1/2 unit or less, and running at $4\Delta x$ 1/4 unit or less. That means that for the same cost as the extra LG or FF run you could do a complete convergence study, which tells you so much more: are you convergent at all, if so are you in the asymptotic region, if so what is the error estimate?

... We dont think that adding these results to the current paper would be appropriate for two reasons. First, the objective of this paper is clearly to revisit the MISMIP and MISMIP3d experiments, see the influence of the three methods on the results and produce new Stokes results for MISMIP3d at higher resolutions.

In that case I don't think that the material is broad or deep enough to merit a full publication (not so concerned about the length – I would endorse a manuscript of this size, or indeed, a single page, if contributed significant new understanding.) The crux here does seem to be a disagreement between the authors and referees on whether comparing LG,FF, and DI tells us very much or not, and I will of course respect the editor's decision on this.

Second, adopting a more realistic friction which vanishes at the GL (as we suggest in the conclusion for the future model intercomparisons) will, by construction, lead to the same results for the three methods, and therefore not being relevant regarding the main purpose of the paper. In conclusion, building new setups with more realistic friction at the GL is indubitably very interesting, but the objective would be then to study if it reduces the difference between the Stokes and the lower order approximations. We are working on that at LGGE, but it is clearly beyond the scope of this paper.

I don't entirely agree with this: I think it would be interesting to see if the Leguy or Tsai laws produced lower numerical error for Elmer/Ice. As above, and in the earlier review, the fact that the LG,FF,DI methods will give the same results is another reason that I don't think the LG/FF/DI comparison actually tells us all that much about the error. I agree that comparison with between Stokes and lower order approximations is clearly out of scope.

The paper has been modified to state more clearly the objective of the presented work. We agree it is short but it however fulfils the TC criteria in terms of length. To make the discussion clearer, we have added two figures and extended some parts.

I note that this paper was originally submitted as a brief communication. Personally, I think that would have been the better choice. I don't want to further delay the publication of the revised MISMIP3d results, but if the authors have firmly rejected the idea of placing these results inside a larger paper of some sort, then I suggest that some of the work could be trimmed to produce a brief communication. I suggest that the important results center on the convergence of the DI method with resolution (ie, taking the DI results from figs 5a,b,c,d to produce a single figure and similarly with (7,8,9). Also it would need to note that the older LG method was used before and gives similar but incrementally worse performance, and a new FF method is much the same.