Response to Referee ♯2

We thank Tamsin Edwards, for a number of detailed review that significantly helped us to improve the quality of our submission. Our full response is as follows.

Congratulations to the authors on a thorough study of an important scientific question. It is a systematic study of a number of important ice sheet modelling uncertainties, mirroring as closely as possible many of the choices made by modellers in the SeaRISE multi-model ensemble (with additional combinations to deconvolve aspects of these). The aim is to quantify which are the most important uncertainties, and therefore to try and understand the source(s) of the large spread in that ensemble. The ice sheet modelling community should be assessing the impacts of initialisation and parameterisation choices systematically, like this, as standard. It is good to apply this to the SeaRISE choices, but the study is also more widely applicable âĂŞ for general considerations of how to do ice sheet modelling, but also for the current question of how to perform the experiments for the first CMIP endorsed ice sheet model intercomparison project, ISMIP6. I heartily agree with the authors statement that "It would be preferable that all participating models perform one common and highly controlled experiment which allows effective identification of the uncertainties due to specific variations in ice-sheet models."
I therefore think the results presented here are an important contribution to the growing literature on this challenging topic.

Thanks a lot for encouraging us. We took all the points you made seriously. Most of them are already included in the manuscript. Revision related to the additinal experiment will be included (in the main text and/or supplementary) in the submission.

I am happy to see the authors test multiple combinations of changes, to check for interactions between the uncertainties. Not every combination is performed, but the simulations presented are substantial.

Actually to test multiple combinations of changes not shown in the present paper is much harder a task than that at a glance. We have performed many additional combinations during the open discussion, which will be inserted in the text, figures and/or supplementary.

The main problem with the presentation of the results is that they over-claim their ability to explain the reasons for the SeaRISE spread. The title should say "potential" sources for spread — and similarily throughout the paper. The study highlights which SeaRISE choices could explain the spread, but due to the limits of using one model and only exploring part of the matrix of choices, they cannnot definitively say if they have found the reasons.

In the present paper, actually we do not intent to identify a definite reason for the SeaRISE spread (in introduction we state that 'This paper does not intend to cover the sensitivities of all of the aspects'). Due to the original title and/or some improper phrasing in the text, we may give to the readers/referees an impression that we solved the SeaRISE spread by comprehensive sensitivity experiments, but, we agree, not true. So, we modify the tones of these phrasing down to

"potential" explanations in the revised manuscript (titles and through the text). Thanks a lot for your comment.

The authors also go part of the way towards defining a useful "benchmark" experiment, but not quite far enough in either the definition or its justification. Is this purely for model intercomparison purposes, or for more realistic projections? If the latter, the authors should comment on the quality of spin-up (see comments: the sum of the squared residuals would be a better assessment of agreement with observed geometry). It would also require other discussion of the quality of the initial state — e.g. realistic velocities, size of drift? If the benchmark is only intended as an abstract test for model intercomparison, this should be made clearer.

No, the former, just for an abstract test as you pointed out. This section is revised to be clearer.

There should be a clear summary of the SeaRISE choices that have not been explored (whether the choices themselves, or particular combinations of them) — how much of the SeaRISE "matrix" have you covered? There should also be a list of the major other modelling choices not tested — i.e. not "SeaRISE" choices but the main things that vary across models such as enhancement factor, method of obtaining basal friction, flow approximation, resolution. These are mentioned briefly in the Conclusion but should also be in the methods and discussion.

Yes, we expand the SeaRISE description in the methods and discussion. Actually we cover only a portion of the SeaRISE matrix, partly because not all the aspects are explicitly documented enough to emulate by IcIES, and partly because it is too much.

Related to this, there is not enough comparison with the dependence of SeaRISE results on these choices. Do the models with free geometry tend to show larger changes than those without, as found in this study? Similarly, do the models that use the Tarasov melting parameterisation tend to show larger changes than those with the Huybrechts? If they do, it confirms the results of this study across multiple models. If not, it would show that (a) the effect of initialisation choices are highly model dependent or, more likely, (b) structural and parametric uncertainties are just as important as initialisation. It would not negate the importance of this study at all - but it would show the situation is worse than we thought! In other words, it would show that if all SeaRISE models repeated this study, the range could widen by a lot more. This would be further support for the authors statement that it is important to systematically control and study uncertainties with designed experiments.

Mostly no. Tarasov melting is only used by IcIES original submission. Huybrechts melting is used by some models, but we are afraid that there are not enough number of models to confirm the results. So we took your latter suggestions in the text, which further encourages our 'benchmark' experiment. Thanks a lot for the suggestion.

Without the above comparisons to SeaRISE, the authors should remove their claims of explaining the SeaRISE spread. Even with them, they should still tone these down to "potential" explanations.

Yes, we have changed the tone as explained already.

Some minor points occur throughout. In terms of clarity, the authors frequently use the broad phrase "initialization methods" when they really refer to a specific aspect of this - e.g. transient vs steady state, or treatment of margin. This should be replaced throughout for clarity. Consistency of referencing needs checking — e.g. Aðalgeirsdóttir et al. (2014) should be mentioned earlier. The authors often use "primary" and "dominant", which imply "the most important of all choices". They should replace these everywhere with "more important than X" for the comparison being made at that point.

We revised the phrasing through the text as you suggested.

This is a difficult study to explain, and an important one to get right. I have therefore included many detailed suggestions to clarify and improve the text.

We really appreciate your careful reading. Most of the detailed suggestions are included as follows.

MAIN SCIENTIFIC POINTS

1389 Need to also say disadvantage of free geometry here — ice sheet too big.

A text is inserted to explain what you suggest. Ice sheet can be too small also, so only mention that it can deviate from the reality.

1396/21 Except the SD, which is larger.

Revised to be more accurate.

/23–27 Why is the bedrock changed only once, not for multiple combinations as for other things? Is this because (a) the effect is small? (b) most models used the later version? (c) it is newer therefore preferred?

All, mainly (a) and (b).

/10 (and also 1389/16) Language needs clarifying: g-ig spin up and tuning/inversion for basal friction are not mutually exclusive, as they are inferring different state variables. E.g. in Edwards et al. (2014), Elmer/Ice used ice temperatures from a SICOPOLIS glacial-interglacial spin-up then performed inversion for basal friction. And clarify "basal sliding enhancement factors" — does this mean basal sliding coefficients and (SIA and SSA) enhancement factors?

The former. Text is inserted to clarify the point. The latter. Corrected to 'basal sliding coefficients' not to confuse.

Table 2 caption: columns are not ratios but percentages.

Corrected accordingly.

Table 2. Given there are compensating errors, please give the sum of the squared residuals with the observed geometry. This will give more information on which is really the best geometry, not only (by accident) the best volume. I think this is crucial information for making preliminary judgements about the most suitable benchmark initialization.

The root mean square of the residuals are appended. Legend and the main text are modified according to the revised table.

1400/18–20 Doesn't it show the third largest, not the largest?!

You are right. Text is corrected.

1402/1–5 "higher climate scenarios" — for v4, I think the absolute increase from A to C is about the same for C2 and C3, right? And the % increase *decreases* from C1 to C3, which is also relevant. Also worth commenting that v4 makes proportionally less difference in higher climate scenarios, i.e. no longer a near linear response to sliding coefficient as it is for C1.

You are right. Explanation is expanded, thanks.

/13 "Even under larger basal sliding coefficients" — why "even"? Wouldn't you expect the change in initialisation to have even more effect for these? (because it removes their strong effect in the free spin-up). And in fact the effect is bigger for v4 than v1 so you could say "Especially for" instead.

Right. Replaced as you suggested.

/18–19 "including the corresponding constant future-climate scenario case C0" — isn't it anomalies with respect to this (as for Fig. 2)?

We are very sorry that the text to describe Fig. 2 was wrong. Your are right. The text is not correct, because C0 is not included in the figure. Text is corrected accordingly, as 'Figure 2 shows the changes in VAF relative to that under the constant climate scenario **C0** obtained by experiments B, F and D, over all the combinations of climate scenarios and sliding coefficients.'

1403/1–5 Isn't it worth commenting on that the effect of the non-equilibrium thermal state (D to F) is nonlinear in basal sliding coefficient and climate? i.e. not systematically smaller or larger? I think for v1, C1 decreases, C2 the same and C3 increases, while for v4 it always increases (i.e. I don't think it's just a function of delta_ VAF)

I assumed you mean B and F (this is not your fault but us, not to name the experiment effectively). Comment is inserted as your suggestion, as 'The effect of the non-equilibrium thermal state is not systemically larger (F to B). In the case of **v1** basal sliding, F shows smaller response than B under the **C1** scenario, similar response under the **C2**, larger response under the **C3**, respectively, while in the case of **v4** all F shows larger under the three scenarios.'

/10–15 I didn't understand this summary — partly because it says observed geometry, when F and F_ s are not fixed to observed. Is it saying that as long as the final geometry is the same, it doesn't make much difference whether the spin-up was transient climate or steady state? Does this mean internal thermal state doesn't matter at all? (i.e. whether it's at equilbrium with geometry, and whether it's derived from 125ka transient or 0ka steady state)? Also — C0 isn't shown.

Yes, that is what we are saying. We revised this part to be clearer. As shown in the second last comment, C0 is not shown because they are all difference relative to C0 results.

1404/14–16 "The retreat of the ice-sheet margin over north-western Greenland is not seen in the B' : v4 cases (Fig. 4f)." I can't see a difference in the north-west between B' v1, v2 and v4, only between B and B'.

Corrected. The retreat is not seen in all the B' cases.

1405/14–23 This has been cut and pasted from p1402 with only minor edits, and should be deleted because it focuses on comparing initialisation methods (rather than SMB parameterisations) with each other. In my view the points of interest left to discuss are F' vs F and F_ s' to F_ s, which look very similar to B' vs B in absolute changes — i.e. the SMB change seems to be related to B0 (large effect) vs observed geometry (small), and not the transient vs steady state, if I have the experiments correct. Possibly the fractional changes are more similar though?

The former, shortened. The latter: Sorry, this comment is also raised by wrong description of Fig. 2. These are not absolute changes but relative changes from constant climate run. The three experiments F, Fs and B start from an identical topography. Also, three experiments F' Fs' B' start from another identical topography. Since the surface mass balance are ultimately a function of the topography only, the former three and the latter each starts from identical surface mass balances.

1406/7–8 It's not useful to say that one thing is the biggest increase and another the biggest decrease, because it depends on the order you do things! Changing from fixed obs to free geometry (D to B), is a much bigger enhancement of response than turning sub-melt sliding onâĂę! (Obviously understanding the direction of response is important, and so is ordering by magnitude, but not ordering by signed magnitude).

Revised to make it clear, as 'Among these four aspects, the inclusion of submelt sliding enhances the ice-sheet response strongest (A to B), but using "fixed-topography" spin-up cancels and even reduces this impact (B to D).'

/10 I disagree that the sensitivity to basal sliding coefficient is a function of delta_ VAF — if that were true, then A and O would always look similar to B, and for example, D' C3 (3 to 3.5) would look similar to B C2 (3–4.5). I think it's a function of free vs fixed geometry (with some enhancement for larger delta_ VAF, but nonlinearly — the C3 range of B v1–v4 is similar to the C2)

You are right. Removed.

"The spread of the results due to different basal sliding coefficients is slightly larger under the C3 scenario." — as above, to me it looks too similar to C2 to comment on.

Corrected as your comment.

1407/5 You can't say you are defining / quantifying the "source of the spread in SeaRISE experiments". You are seeing if you can reproduce the magnitude of variation in delta_ VAF (N.B. not model response, which would include spatial elevation patterns, velocities etc) by changing a subset of the possible configurations (i.e. not all combinations of all parameters/choices that are changed across SeaRISE models) in one model (not all possible model structures). This is a very interesting study, but it is induction not deductionâĂę Your results point to potential sources of variation, but cannot definitively identify them.

Yes, we agree. This part no more exists according to other modification.

/12 Is there no drift at all for free spin up? I would have thought model and obs (eg bedrock, GHF) errors would still give some.

Yes, we do have some, but the drift of free spin-up run is not 'artificial'. Maybe this term is not good because any simulation is more ore less artificial. Rewritten this part without 'artificial'.

1408/4 Can you please explain this in more detail, as I don't understand what is done.

This part is revised to: "This 14.5cm effect is about 11 % of the simulated VAF response obtained by D C3:v4 case, and thus the effect of the internal non-equilibrium state is expected to be kept minor than the total sensitivity."

/5 "Therefore, future-climate experiments...preferred..." — what is the basis for this conclusion? (The volume of the A v4 spin-up is the closest to the observed!). What other information do you have that fixed geometry is better? That it's less sensitive to other perturbations than the free? How do you know that's desirable? Is it because it's easier to spin-up with steady state than transient? It's not clear what the justification is.

Explanation is inserted, as "In other words, initial topography has more control on the future projection, *in terms of relative to constant scenario runs*, than initial internal temperature field." The point is that, if two experiments starts from the identical topography, one with consistent temperature field (free spin-up), the other with inconsistent field (fixed spin-up), and all other model configuration are the same between two, then that simulated volume responses to future warming scenario *relative to those to constant scenario* are similar during five-hundred years.

/10–15 I don't understand this comparison. Free vs fixed geometry (B–D) is a bigger effect than either of these, so isn't that the "primary source of uncertainties in the simulated short-term future projections of the Greenland ice-sheet". And better to say 500 year time scale, as those focused on 1–200 years won't see this as short- term.

We agree, not the primary. The text is revised.

1409/5 Please give the total trend for each run after 500 years — at minimum the range and which are the largest drift runs. It would be useful to separate this for transient and steady state climate runs.

The ranges of the trend (by C0) are inserted: "Simulated trends vary among the configuration: ranges from −45cm E-v4) to +24cm (D'-v1) after 500 years among transient experiments and −21cm (Es-v4) to +24cm (Es'-v1) among steady-state experiments, respectively."

/8 "The present paper concludes that such long-term memory has a smaller impact" — this isn't shown

This is concluded by the results that the transient spin-up (which has memory of long-term changes in climate) and the steady-state spin-up (which does not) are almost same. Text is revised to be clear, as 'It is expected that such long-term memory has a smaller impact for the future *changes* in ice-sheet volume at least during next 500 years, compared with the changes due to future surface climate scenarios, because the results of transient spin-up (with long-term memory) and steady-state spin-up (without) show similar responses,'

Discussion section generally:

Can you give an ordered list (or figure) of the one-at-a-time effects? I think this would bring clarity to the multitude of bars... i.e. for a given v and C, and only for pairs of simulations where one thing is changed, what are the largest to smallest effects of changing that one thing? Fraction might be more useful than absolute. So 15 pairs, I think — which could be repeated for v1:C1, v:C3, v4:C1, v4:C3? (in Supplementary Material), order by magnitude the fractional changes of:

```
A--O
B--A
D--B
E--D
F--D
D_s -- D
D_s' -- D_s
F_s -- F
F_s' -- F_s
B'--B
D'--D
D_s'--D_s
E'--E
F'--F
F_s'--F_s
```

Or could just show the top 5 out of 15 in each case, maybe.

Sure. We will add such figure or table to supplementary. Two comparison are duplicated in the list above so there are 13 pairs totally. The following is a draft of the ordered table for the four combination of C-v pairs. Fractional changes (0 means no change) are shown.

C1 v1		C3 v1		C1 v4		C3 v4	
E′-E	-0.305	F-D	+0.398	F-D	+0.807	F-D	+0.452
D-B	-0.287	E′-E	-0.370	D-B	-0.436	E′-E	-0.409
F-D	+0.246	F'_s-F_s	-0.361	F'_s-F_s	-0.433	F′-F	-0.363
D'_s-D_s	-0.245	F′-F	-0.359	F′-F	-0.425	F'_s-F_s	-0.359
F'_s-F_s	-0.244	B′-B	-0.337	B′-B	-0.386	D'_s-D_s	-0.353
E-D	-0.239	D′-D	-0.335	B-A	+0.379	D′-D	-0.352
F′-F	-0.235	D'_s-D_s	-0.333	E′-E	-0.259	B′-B	-0.347
D′-D	-0.234	D-B	-0.261	E-D	-0.243	D-B	-0.257
B′-B	-0.228	E-D	-0.109	D'_s-D_s	-0.188	B-A	+0.155
B-A	+0.034	B-A	+0.045	D′-D	-0.180	E-D	-0.085
A-O	+0.034	F_s-F	+0.014	F_s-F	+0.015	A-O	+0.009
D_s-D	+0.021	A-O	+0.011	A-O	+0.013	D_s-D	+0.007
F_s-F	+0.013	D_s-D	+0.009	D_s-D	+0.007	F_s-F	+0.004

I would like to see more about why the configurations produce the response they do, and if they are expected — there is some discussion about increasing flow to margins, increasing sensitivity to climate etc — but I think it should be added for each result, even if the reason seems obvious to you.

Such discussion are inserted for each result. This is partly explained by initial condition and partly explained by initial drifts.

/22 "The results show that the main sources of the spread in the SeaRISE experiments" → no, in this particular set of experiments reproducing possible SeaRISE choices

This part is revised to use the term 'potential'.

/24 "As already proposed in the SeaRISE papers, and confirmed quantitatively in the present paper, the impacts of these two aspects are of comparable magnitude." — again you need to distinguish these two related but not comparable studies.

The text is deleted.

1410/2–3 "temperature is allowed to be evolved according to the surface temperature historyâĂęevaluated" — why is this preferred over steady state? It wasn't really evaluated either.

We agree that this is not preferred over steady state. The text is revised.

/9 "consistent or inconsistent." I don't think this is tested as it requires uncertainty estimates. "Similar or dissimilar" would be OK.

Right. Corrected as you suggested.

/21 "Here we propose a model intercomparison study " — all of this is new so shouldn't be in the Conclusion but in the Discussion.

This part is completely moved to the bottom of Discussion.

1411/1–12 Why choose no advance — why do you think it is better? (it might be, but needs justification). Or is it no better, you just pick it arbitrarily (also OK! but say so).

We cannot say it is either better or not. One thing we expect they choose no advance is from a technical reason: we do not have much constraint for surface melting and/or surface velocity over ice-free grids. If advance is freely allowed, we have to define for example the basal sliding coefficients over those grids, which may have much degree of freedom. Also, we expect some model have a difficulty, although possible, in regeneration of grid system if ice-sheet area is expanded. Such explanation is inserted.

Need to be clearer why benchmark includes either method — "very heavy full-stokes" should be "computationally expensive Full Stokes".

Explanation inserted.

/14–18 "entirely" — No! Also resolution, other parameters? Treatment of calving, retreat allowed Are you confident in recommending a benchmark that appears to systematically reduce response? (fixed obs geometry, no advance) Is this purely for model intercomparison or for more realistic projections?

The latter: this is purely model intercomparison, not for realistic projections. Still many modeling studies (as SeaRISE) separate ice-sheet modeling and other modeling (climate, ocean,...). Realistic projections require realistic scenarios which are also uncertain. SeaRISE presented that even if the 'scenarios' are the same there are variations in the simulated response of Greenland ice sheet. This paper presented, however, that the variations are partly from the structural differences which can be avoided. In other words, uncertainties due to all the aspects except for the surface mass balance (not 'temperature' scenarios) are expected to be extracted. We do not think the benchmark reduce the response. This is just for evaluation of the ice-sheet model uncertainties to the same forcing. The former: other parameters you mentioned are also included in 'entirely'. This part was revised to be clearer.

Where is the SMB from (as above)? Why look at uncertainty due to PDD scheme if all models use the same SMB? (maybe this is clear once we know what the SMB is)

We have no idea what is the best SMB. Although we agree it is another option to use the identical PDD scheme, it is easier to use the SMB field computed externally in advance than to that computed internally in particular for the purpose of the benchmark. The reason is we have to provide not the equations but the subroutine in order to make the SMB really identical among the models, probably in two or more languages. An explanation is inserted as another options, as "or provide an identical surface mass balance *subroutine* (not a scheme to keep it really identical among the models) as well as scenarios,"

CLARIFICATIONS AND SMALLER CHANGES

1386/19 "changes, for example, " → "changes such as"

Corrected accordingly.

1388/20 Needs a bit more info to explain.

More information is added: "One method applies a Heaviside function at the pressure-melting point of the basal temperature, i.e., the basal sliding is set to zero when below the pressure-melting point. Others apply a smooth sliding transition around the pressure-melting point (Hindmarsh and Le Meur, 2001), i.e., the basal sliding gradually becomes close to zero below the pressure-melting point, partly for numerical stability and partly for physical reasons to introduce sub-grid scale variation of the basal sliding."

1389/12 What? I think delete "only"

Deleted.

/14–28. Confusing structure. Clearer if you write: 19 "The second and third methods are..." and 21–22 "The first of these, "free spinning-up",..."

Revised as you suggested.

1390/24 Isn't the "flux-corrected" the same as the "synthetic corrections" in line 10–11 on the same page?

Yes, they are the same in principle, while different in detail. An explanation is inserted.

/28 Yes it has, e.g. by Goelzer et al. (2013)

Goelzer et al. is mentioned around here.

1391/12 Add "only" before "retreat is allowed"

Inserted.

/16–21 Tech vs physical are not great terms. Is it internal/structural vs inputs? OR dynamics vs SMB? Also "indirect processes" vs "direct impact of the model to response" is unclear; and 19 needs comma after configuration.

Revised.

1392/3 brackets wrong for ref

Corrected.

/6 SD - define - standard deviation? and "short-term statistical air temperature fluctuation" is unclear — is it e.g. Gaussian noise added to parameterised monthly data?

Right, defined.

/8 ^the^ GrIS

Inserted.

/10 diversion → divergence / dispersion?

All the term 'diversion' are completely replaced 'dispersion'.

/11 qualitatively → quantitatively

Corrected.

1394/3 Not clear. What is the background temperature field - annual data or annual and summer? Does computation mean lapse rate correction only, or also the interpolation to obtain monthly data? And is this a spatially-varying lapse rate correction?

This part is revised to be clearer. Monthly temperature is computed by an interpolation of two independent parameterization of the annual and summer temperature. The lapse rates are spatially constant.

/15 "obtained by" → estimated from

Revised.

1395/14 ^the mean of^?

Inserted.

1396/23–27 Need to mention all experiments (e.g. D_s, F) here. Why no C?

Description of all the experiments are inserted. C is avoided because it is used to specify scenarioes C0, C1, C2 and C3.

For a big improvement in readability, name experiments using most important change/comparisons, e.g.:
A = NO_SLIDE, B = TRANS_FREE, D = TRANS_FIX, E = TRANS_FIX_NOADV, D_s = SS, E_s = SS_NOADV, F = TRANS_FIXFREE, F_S = SS_FIXFREE
Then use the primes as before.

There are already not a few comparison patterns among the experiment in this manuscript. Moreover, since you suggest to test multiple combinations of changes not presented in this manuscript, we are afraid that such changes in the naming convention rather confuse the readers. However we agree the manuscript need improvement in readability more or less, so we will try several patterns in the revised manuscript and will choose the optimum pattern. At the moment candidates are (i) keep the name, but every now and then the most important changes are added in the text, as 'A (O/jhkp)' 'A (B/no submelt)' (ii) rename all the experiments with four symbols, such as A to AFNJ (Advance; Free; submelt No; JHKP), D to AOYJ (Advance, Obs-fixed, submelt Yes; JHKP), E to ROYJ (Retreat, Obs-fixed, submelt Yes; JHKP), F to ATYJ (Advance, Transient-fixed; ...) and so on.

I find the sliding parts a bit confusing to read at first. It feels like more things are being discussed than are! It would be clearer to spell out that the Heaviside step function prevents submelt sliding, the exponential allows it, otherwise it sounds at first reading like two things are varied. And that the v2 and v4 are testing the model tuning, while the SeaRISE experiments vary the same thing but test the effect of increased sliding from climatic forcing. And finally some discussion of how these things interact (basal sliding parameterisation and value(s) of basal friction coefficient).

This part is revised as your suggestion.

1397/ Suggest for clarity changing "In the first method, similar to the free spinning-up, a steady-state simulation is performed under the climate field at 125 ka with fixed ice-sheet and bedrock topography of the present-day state and only the temperature can evolve. Subsequently, the climate history from 125 to 0 ka is used to force the internal ice-sheet temperature." to: 'The first is identical to the free spinning-up except that the ice sheet and bedrock topographies are fixed to the present day state and only the temperature can evolve". I would also delete "Experiment D uses the same configuration as B except for using this fixed topography transient spinning-up." and similarly line 14–15.

Revised and deleted.

1398/19–21 Delete (repetition).

Deleted.

1399/3–8 Seeing as the SD is 5.5K in all experiments, why not just mention it at 1396/21 and not here? i.e. it's not part of the experiment.

This part is moved to the point you suggested.

/9–11 Delete? — unnecessary.

Deleted.

/20–24 is confusing. Does "final state" mean just ice temperatures or everything? Does "is adopted for" mean at the start, then spun up? Or used at the end somehow? e.g. temperatures combined with other (non-temperature) fields?

We agree it confusing a lot. This part is revised.

/25 Suggest editing for clarity: replace "To evaluate... respectively." with "To evaluate...., ^further^ fixed topography ... (experiments F_s and F_s'). Instead of the topography being fixed at the present-day observation, as for D_s and D_s', it is fixed at the final topography of the spin-up phases of experiments B and BâĂš, respectively."

Revised ad suggested with slight modification.

Section 4.6 is generally tricky so a bulleted list would be useful. Something like this?

```
. B free topography + transient temperature
. F fix to free topography + transient temperature
. D fix to observed topography + steady state temperature
. F_s fix to free topography + steady state temperature
```

I know you have Table 1, but a summary in the main text of how these four relate to each other would help.

A list is appended as you suggested.

Table 1 and the main text needs to say that E" is the benchmark experiment in the Appendix, and that it uses a different SMB, and what that is. Otherwise the reader is left hanging wondering what it is.

Description is inserted.

I'd also suggest putting all the prime (') experiments together at the end, as it's harder to keep track of the other configurations. And could the differences from the previous row be put in bold? e.g.

```
O
A JHKP...
B ... sub-melt ...
D ... fixed obs geom ...
E ... no advance
F ... fixed free geom, free margin
```

Then to simplify the table you could just say in the caption that _ s means steady state initialization, while prime means switching from T to H melt âĂŞ a bit like modifiers for katakana and hiragana :) The v1,2,4 and C1, 2, 3 could be in the caption too.

Tables and captions are revised accordingly.

Table 2 would benefit from a brief recap of v1, v2 and v4 in the caption (as you do for Figure 2).

Descriptions are inserted.

Figure 1. Can you make each experiment name the same colour as the line? In the caption, I think it would be clearer to write "...obtained by future climate runs under C1 (A1B climate forcing) climate scenario, with "standard" sliding coefficient (v1), in terms of the differenceâĂę Each line is a different experimental configuration... "

Good idea. Thanks. Experiment names are also colored. Caption is revised.

1400/5 This needs a pointer that other results will be described later, and a clearer separation between discussion of O and others. e.g. move "Simulated responses become larger with enhanced basal sliding coefficient" to 5.2.

Such pointer is inserted. The latter text is kept because it is still the results of O, not B.

/19 "shows"

This part is revised already.

Figure 2. Again, I think it's clearer to mention the 3 climate scenarios first, then the experimental configurations. I really like having both VAF and SLE scales, but I think it would be better to put the SLE on the right y axis (with common major grid lines, as they are for C1 and C3).

Caption is reordered. Axis of the figures are modified accordingly.

1401/3 Is it +17 cm increase from v1 to v4? Current wording sounds like absolute for v4.

The text is revised to be clear, as "17 cm more than the upper boundary of the original range"

Figure 4 — Show observed topography too, or show all as anomalies with respect to this, so that differences can be seen clearly.

Difference in the topography from the present-day observation will be included in the supplementary.

/6 → "Simulated VAF responses are ^^ affected by replacing the bed topography of a few regions, ^but^ are less..."

Revised accordingly.

/20 Not very clear. Suggest: "For O and A, quadrupling the basal sliding coefficient varies the volume by around 5% of observed, but for B by more than 12%."

Revised.

Figs 3 and 4 appear in the opposite order in the text.

Actually, Fig. 3 is already mentioned (at 1400/21) before the first appearance of Fig. 4 (1401/22). However, since Fig. 3 is not discussed until 1402/16, we exchanged the order of two figures, and modified main text accordingly. Thanks.

Fig 4 Add "and sub-melt sliding", as O and A are also free.

Inserted.

1402/1 greater → more. Add "(comparing B with A)"

Corrected and appended.

/3 missing bracket

Inserted.

/4 in the VAF → between B and A

Revised.

/9 "This is large enough to cancel" → "This more than cancels"

Revised.

/12 "initialization method" is too broad a term — replace with "whether the geometry is free or fixed to observed" or similar.

Revised accordingly.

/25 "Of all the combinations" Should be "Over all.." and ideally also add "of climate and sliding coefficient" to be clearer.

Modified and inserted.

/28 "the the"

Corrected.

1403/1 Again "initialization methods" is too broad — replace with free vs fixed geometry. Also: "when they ^are^ evaluated".

Corrected.

/3 "Through the elevation-ablation feedback, the impact of the non-equilibrium thermal state is larger in cases of higher sensitivity." — please expand/rephrase, as it's not clear what it means. Is it "The effect of the non-equilibrium thermal state is larger for larger delta_ VAF, because the elevation-ablation feedback amplifies the geometry changes"? (but see comment for 1403/1–5 — the picture is not that clear).

Right. Revised as you suggested.

/6 I'd argue "and B" should be deleted, as you've just compared B and F as being different methods even if the results are very similar. Or should this read "F_ s and D_ s are equivalent to F and D"?

Deleted as you suggested.

/20–22 "Thus, under mild climate warming scenarios like C1, the choice of initialization method and the margin treatment has dominant effect on the response of Greenland ice-sheet over 500 years." — what does this mean — compared with the bedrock, sub-melt sliding and basal sliding coefficient? The effect of the margin is smaller than that of the coefficient, at least for some runs. And the SMB has not yet been discussed.

We compared with bedrock and sub-melt sliding. Text is revised.

1404/21, /25 and /27: "initialization methods" → free vs fixed as before

Modified accordingly.

/26 ambiguous — assume these numbers are for B to B' not B to D? Better to put them earlier in the sentence.

Right. Moved to earlier.

1405/4 — I think "large" is better than "dominant", or "similarly large influences... as"

We modified with the latter suggestion.

/5, /9 and /21 initialization methods → free vs fixed

Modified accordingly.

/12 60% is indeed a big reduction — so to be fair I think you should also say in Section 5.3 that B to D reduces to around 50% or less.

Inserted accordingly in the section.

/26 different initialization → different fixed geometry

Revised accordingly.

/27 Repetition of first line of paragraph. What is the difference? Looks to me like SMB is more important than which geometry is fixed. You go on to say "Further, of all.." but this should be "Over allâĂę", because it is not "further", it is the point of the paragraph. And if you want to mention steady state, this needs to be in the paragraph start. I think this means comparing not only B' and F with B, but also F_s with B? or maybe with B_s? This paragraph needs expanding.

Right. The repetition is revised and the steady-state is mentioned at the beginning accordingly. Thanks.

1406/4 Why isn't F included in this list? Are the "_s" experiments included? What about v1, 2 and v4? Is it everything except C1–3 and SMB? I think there needs to be greater clarity. I'd say SMB parameterisation was technical, but I agree it's helpful to think of it separately from the ice flow.

F is excluded here because of its characteristic: it is an additional configuration just for comparison between B and D. This part is revised to mention v1–4 and C1–3 accordingly. Also, the word 'technical' is removed.

/19–22 This sentence is confusing and a bit vague. Suggest e.g. "Thus the source of spread in SeaRISE experiments can only partly be explained by variations in the experimental configuration of technical aspects ^of ice flow^." and then something more precise: "The most influential of these is the specification of free or fixed geometry".

Revised accordingly.

1407/5 This makes it sound like a third thing was identified as the primary source of spread by SeaRISE. I think you mean something like: "In the series of the experiments in the present paper, the choices that have greatest effect on the response are the method to compute the surface mass balance, and the way to initialize the ice-sheet, which have comparable effect. This is consistent with the discussion [speculation?!] of the possible reasons for spread in the SeaRISE results by Bindschadler et al. (2013) and Nowicki et al. (2013)."

Right. Modified accordingly.

/7 "may have a certain" → has some

Revised.

/14–27 I feel this is more repetitionâĂę Reword/focus? And need "Over all the" again.

Revised to be focused, accordingly.

1408/11 Missing full stop.

Inserted.

/21 " it may be better" again needs justification

This part is deleted.

/24 Needs to be clearer whether this is all for transient climate (i.e. only free vs fixed geometry), or if it also apples to / means steady state climate too.

Both. Modified accordingly.

/26 "attributed to the difference in the application of the technical methods such as initialization and free evolving margin, and the difference in the surface melting parameterization." — difference with what? Be more specific, e.g. "attributed to the use of a free geometry and margin during spin up"

Revised to be more specific, as 'the use of a free topography during the spin-up, free evolving margin during the future experiment, and the difference in the surface melting parameterization.'

/6 "showed the divergence or convergence of..." → something like "showed the degree to which current ice sheet models and modelling choices diverge".

Revised accordingly.

/20 "but" → "except", otherwise it sounds like we did do it!

Corrected.

1409/18 identify → "try and understand" or similar. Similarly at 1387/25

Revised at the two places.

1411/1–12 This needs editing and expanding. I'd say temperature is climate, SMB is not. Do you mean SMB from an uncoupled PDD scheme with topography feedback switched off, or SMB from a regional climate model?

Either PDD or regional climate model is all right, as far as identical among the models. This part is revised.

"are less controlled except" → "not specified, except"

Replaced.

/14–18 Why choose configuration E'?

Because among the configuration in this paper this is the closest to the one of ISSM.

"future surface mass balance is imposed using the SeaRISE datasets without any correction." Which SMB dataset? (C1 as below?) Correction for what? (topography, albedo, both?)

Although most of the models did not use, SeaRISE provided a transient future scenario of SMB computed by a method not explicitly described in the SeaRISE papers. We revised this part, as: "The parameters of the PDD is described on `http://websrv.cs.umt.edu/isis/index.php/Future_Climate_Data`, where the standard deviation of the short-term statistical air temperature fluctuations is set as 4.5 K, the PDD factor are set as 3 and 8 mm ice equivalent per day per degree for snow and ice, respectively."

Contradicts "surface mass balance is imposed ^on^ the SeaRISE datasets without any correction;" → "with?

Right. Corrected.

Fixed calving front — is this the same as no margin advance?

This is what is written in Tab. 1 of the SeaRISE paper. We expect that not only advance but also retreat are prohibited. Explanation is inserted.

List similarities first, then differences.

Reordered accordingly.

Assuming sub-melt sliding same?

There are no information about sub-melt sliding. Explanation is inserted.

Anisotropic mesh is a bit jargon-y/unclear — ideally describe it.

Revised to 'numerical grid system'.

Figure A1 — reduce scale to 1.6 min VAF

Rescaled accordingly.