Summary of major changes

Figure 1 and 2. The *m* and *r* have been added to each panel, in addition RACMO2.3 and CESM1.0 in text are annotated to the first figure in colors corresponding to their scatter and lines. Y-axis is changed to show the same range for the same units.
Figure 2. Added an additional panel containing snowfall.
Figure 3. Added an additional row with SMB vs elevation for each of the lapse rates.
Figure 4. Added a spatial map of the RACMO2.3 reference data. Changed panel (b) to show EC-1K minus EC-6K.
Supplementary figure. A supplementary figure similar to Figure 1 has been added showing the incoming and outgoing solar and longwave components.

Reviewers comments in black
Author's response in red

Reviewer #1:

General comments

This paper describes an analysis of the online climate downscaling scheme used over the Greenland ice sheet in an older version of CESM, alongside a limited study of how sensitive it is to the temperature lapse rate specified, one of the scheme's key parameters. The topic is timely, although I don't think it's totally clear that this belongs in The Cryosphere rather than Geosci. Model Dev., seeing as it doesn't purport to research anything about the real world, rather evaluate the emergent behaviour of a model parameterisation. That's not meant to imply that I don't think it's an important subject, and it is valuable to highlight these results to a wider audience than those who work on the development of climate models, as it directly bears on the interpretation of model results that are used widely in the cryospheric community. In general I liked it, and as someone working in this area, I found it practically useful.The paper is well structured and clearly written - my main recommendation for an improvement would simply be to show more figures.

We thank the reviewer for this feedback, and we will expand on the figures in the manuscript, in addition to adding more figures to the supplementary material. We submitted this manuscript to TCD instead of GMD, as the study is more focused on physical analysis of surface fluxes and climate impacts resulting from the method.

One might suggest a number of improvements to the downscaling scheme itself, of course, but those would be outside the scope of the work actually presented here.

Detail

title: the editor's initial comments have already touched on the title, but I'm not convinced the current title is as clear as it could be. The quantities being actively downscaled are "climate" variables, not the SMB itself.

The authors' replies to this comment from the editor also say they'd prefer to leave "Greenland" out of the title, as the scheme is general. Personally I'd put it in. The other obvious ice sheet application for this is for Antarctica, but circumstances there are rather different. Sub-gridscale variation in SMB components there is more dominated by dynamic weather considerations rather than pure elevation, and temperatures are such that the lower, melt/bare ice albedoes - the only means by which sub-grid variation in shortwave radiation can really enter in this scheme - should play a much smaller role. That being the case, the analysis and component gradients here probably *are* only really applicable to Greenland. Additionally, no comment is made of how the scheme might perform for other ice sheets - perhaps if the authors wanted to leave the title as it is they could include some discussion about how the scheme might be expected to perform on Antarctica, or if applied to paleo ice sheets in other regions?

If they wish to keep the scope to just modern-day Greenland, how about "Downscaling climate through elevation classes for Greenland ice sheet surface mass balance in an ESM: analysis [etc...]"?

We consider we are downscaling the SMB and surface energy fluxes and not the climate, as the downscaling is done in the land component of the CESM as opposed to e.g., grid refinement in the atmospheric component or statistical downscaling of climate variables.

We are aware of the current application of the EC method to some other ice sheets, but we prefer not to speculate about how adequate it would be for those. Instead, we think future work can use our study as to guide the evaluation to other (paleo) ice sheets.

Pending the editor opinion on this, we could change to "Surface mass balance downscaling through elevation classes in an Earth System Model: application to the Greenland ice sheet".

page 1, line 5: it would be clearer if you note that RACMO is an RCM

Will change to "from the regional climate model (RCM) RACMO2.3"

p1,l20: "leading" would be better as "which would lead"

Will be changed accordingly

p1,l21: "is losing" would be better as "has lost" if you start the sentence with "Since"

Will be changed accordingly

p2,l6: I'd say "ESM" deserved a wider definition than 'a climate model with a carbon cycle'. There are many possible physical components in an ESM, and for certain applications I don't think you would necessarily have to have the carbon cycle part active to still call the model an ESM

Will be changed to something along the lines of: "... (ESMs; coupled climate models capable of simulating the Earth's physical, chemical, and biological processes)".

p2,l8: does the "SMB" contraction need defining in the Introduction proper rather than just in the abstract, which can sometimes stand alone from the main paper?

SMB will also be defined there now.

p2,l10-25: I didn't think the distinction between methods 2. and 3. was terribly clear, or that the "hybrid" variant used by CESM doesn't really sit within method 2. or 3. The section also suggests it's going to list "state of the art downscaling techniques" in general, but this is a wide field and this list seems far from comprehensive - pattern scaling, EOF methods etc

This is a good point, and it will be too much to go through all of the state of the art downscaling methods. Therefore it will be changed to "Most common downscaling techniques for the GrIS SMB are".

The distinction between method 2 and 3 is that method 2 only allows for applying statistical corrections to output of a model. This is, as you mention, a very large field containing EOF methods and pattern scaling, in addition to the rapidly growing toolbox of machine learning techniques. However, for a method to qualify for the hybrid approach the SEB/SMB needs to be explicitly calculated after some statistical downscaling of the atmospheric variables involved.

p3,l1: since CESM1 was superseded by version 2 more than a year ago, I think that somewhere in the introduction it would help if you explicitly noted that you're not using the current release version of CESM, and said why. Perhaps in the Discussion you could also note what, if anything, readers might expect to be different in CESM2, based on what you've learnt and what has changed in the model in the meantime

This study was based on CESM1.0 as the elevation classes were first introduced in this model version. We are preparing a short follow-up study using CESM2. CESM2 uses some of the recommendations made in the text (e.g., a lapse rate for incoming longwave radiation, a lower ice-albedo, and precipitation phase corrections based on surface temperatures).

p4,l24: Why did you use a "minimal", 1K/km lapse rate as a control rather than 0K/km which would effectively deactivate the scheme properly and revert to the type of behaviour seen in most ESMs?

We feared that using a lapse rate of 0 K km^{-1} could have unwanted consequences: the model was designed to have activate elevation class with a certain lapse rate, so completely disabling this feature would potentially lead to model artefacts.

p5,l4: I think it's noted later in the analysis, but you're effectively comparing two (likely completely different) realisations of climate variability during a specific historical period by using an ERA-forced RCM vs the GCM. I think it's worth flagging this up, and anticipating the possible impacts here already.

We will add some sentences to section 2.3 to highlight this difference earlier in the manuscript.

p5,l9 The framework used from here on does rely rather on fitting simple linear relationships to scatter plots of " vs elevation" from all of Greenland. The apparently wide scatter in the figures often make it look like such a simple relationship really isn't a good way to approach the RACMO data being compared with, although the r values given look higher than the scatter shown in the plot might suggest, so perhaps this is more a presentational issue? Since a universal linear gradient is the paradigm being used in CESM - and the CESM fits do often *look* much more linear - it's not an unjustified way to proceed, but some kind of cautionary note should be put in here that this is a potentially over-simplistic way of looking at regionally heterogeneous data from a much higher resolution study, and that for some variables the scatter makes the fits and gradients reported perhaps more qualitative than quantitative.

This is a very valid point. We believe the scatter itself might appear as larger than it statistically is, especially for RACMO, and the reason for the r-values to look higher than the scatter shown is due to the very high number of points in this plot (13,311 for RACMO and 1,551 for CESM as stated in p5,l17). Also, the temperature forcing itself is a linear regression onto the elevation, which makes seeing how (non-)linear different quantities respond to such a forcing an interesting point in itself through the linear correlation with elevation.

p5,l17: I think some 2D plots so that readers can see the regional differences between the RACMO2.3 reference fields you use in this study and your CESM1 SMB would be very useful here. The choice of which figures to put in the main paper in which should be supplementary material will need to be thought about, but in general I think this is the first of a couple of areas in this paper where it would just be useful to be able to see more information than is currently there.

We will add a map of the RACMO2.3 data to Fig. 4, and made EC-1K to an anomaly map wrt. EC-6K.

p7,l29: it wasn't immediately obvious to me why the subset of fluxes shown in figure 3 were the "most relevant"

It is because we focus on the fluxes that are controlling the downscaling: the turbulent fluxes (sensible), net shortwave (albedo), and melt energy. We will also add a fourth row with SMB in the final version, as this might be the most relevant of all to show here.

p7,l33: the CESM albedo does not look very sensitive to the lapse rate - probably worth noting that even at the maximum lapse rate you don't even get to half of the RACMO value.

The numerical difference in albedo gradient is not very large indeed. However, we feel that it is still sensitive to the forcing lapse rate, as even a small change in the albedo gradient leads to a much larger change in SW_{net} *gradients. We will add a statement of how it compares to the RACMO values according to the suggestion.*

p8,l9: why not actually do the SMB scatter plots and show the gradients? Surely they're important enough to show explicitly?

Fig. 3 will be updated to also show the SMB scatter plots.

p8,l31: I really didn't understand the description of the prognostic temperature, or how it was calculated

The prognostic temperature is calculated at each EC in CLM as a result of the calculated energy fluxes and exchanges. Therefore, it is different from the forcing temperature.

p9,l9: I still don't understand why you used a 1K/km experiment as the control, rather than 0K/km?

Please see answer before.

p9,l14: it's not clear in which topography the "mean elevation is lower"

In this part, we mean that the atmospheric grid cell elevation is lower than the land grid cell elevation, even though both are on a 1º grid cell as the atmospheric model requires a more smoothed topography to not force the highest wavenumbers that can cause noise. We will add to the methods section a more clarifying description of the different topographies that goes into the simulation.

p9,l29: the Supplementary info figure is labelled A1 here, but as S1 in one of the links I was given to download

It appears as S1 in my PDF. It was labelled A1 during the initial submission, but was corrected and now appears as S1 in the public discussion paper.

p9,l30: the large discrepancies in the comparison with the reanalysis may be a place where the fact that the reanalysis and the GCM will have different realisations of internal climate variability really plays a role

This is true, we will highlight this more.

p10,l29: as previously noted, I think it's worth flagging up differences between your CESM1 and the new CESM2, which is the version new users will likely pick up. Can you say which, if any, of the recommendations you make have been implemented in the current CESM?

Of the recommendations made in this paper, the following were implemented in CESM2:

- *Lowered albedo*
- *Downscaling of incoming longwave radiation with a fixed elevation dependent gradient*
- *Downscaling of precipitation phase based on surface temperature*
- *More advanced firn simulation*

We decided not to include references to CESM2, both to avoid confusing reader and as the effects of these different parameterizations have on the downscaled SMB is not yet documented.

p11,l20: "certain lapse rates score better for some metrics than others" is a little disingenuous, really. You've done a great job of showing that that the components being directly downscaled via the lapse rates generally cannot be made to match the physical elevation gradients for any value of the lapse rate, and that the final SMB you get only scores well because of fortunate cancellation of these significant errors. At this point, the "lapse rate" you specify almost loses a physical meaning - it's no longer a parameter you might desire to constrain directly through observations to match reality, rather a model control you can tune directly to get the final (SMB) result you want without worrying about the fidelity of the underlying components that go into that result. Something along these lines should be noted in this paragraph, I think

Yes, we agree in that the lapse rate is the tuning control to redistribute energy within a grid cell. We will remove this statement ("certain lapse rates score better …").

p11,l24: implies that it's hard to distinguish between the EC6K and EC9.8K SMB gradients, yet two sentences before states that EC6 has a better SMB gradient but EC9 has the best melt. I'm confused as to whether you can really make a robust distinction between the SMB gradients in the two cases - especially since the SMB vs height scatter plots are not shown for cases other than EC6. Does the r value on the SMB gradient actually justify distinguishing between the two cases? If, in fact, you're only basing that recommendation on the top line of total GrIS SMB in

Table 2, given the size of the standard deviation on the RACMO numbers it would seem difficult to justify saying one is better than the other.

We will add an additional row to Fig. 3 with SMB gradients which clearly show that the SMB gradient in EC-9.8K is very steep compared to RACMO.

Figures

On the whole I feel that the paper could be usefully improved by tweaking the presentation of the figures. Above I've noted that it would be good if 2d plots of the EC6k vs the RACMO2 reference data could be shown, and the actual SMB scatter plots and fits for EC1, EC4.5, EC6 and EC9.8 rather than only summarising this data in a table. It may be that the authors or editor take a view on which figures belong in the main body of the paper and which in Supplementary information, but I do think it would be uesful to show them.

We have followed the recommendations of the reviewer and will add the SMB scatterplots to Fig. 3.

Of all the panels of scatter plots, only Figure 3 includes the useful gradient and r values on the scatter plots themselves - it would be useful if Figures 1 and 2 could show this information as well. In Figure 4, why is the absolute value of SMB shown for the EC1K experiment rather than the more useful difference from EC6K, which is how the information for the EC4K and EC9.8K experiments is shown in panels c) and d)?

We follow the reviewers recommendation of adding the m and r values directly to the plots.