13 Nov 2020

re: "Review of Schneider et al.", Anonymous Referee # 3, 16 Sep 2020

We appreciate you taking the time to express your concerns regarding our manuscript. We are taking them into great consideration as we revise. Following below is our response (in blue) to your enumerated concerns and specific comments, which are italicized here for reference:

1. *The first concern is about the statistical modelling approach explained in Sect 3.2. I didn't get very warm feelings about this. For instance, in the direct comparisons carried out in Sect. 4.1.1 and Figures 2-4, output from a coarse resolution ELM simulation with 6-hourly CRU-NCEP forcing is compared to steady-state profiles from the Herron & Langway model with idealized (synthetic) forcing. To me this feels like comparing apples to pears. Unnecessarily, it seems, since the ELM (like CLM) can also be forced with synthetic data in single-column mode, offering a more direct comparison. After all, to quote Arthern et al. (2010), "Changes in weather and climate can cause temperature, accumulation rate, and depositional density to vary. Consequently, and in violation of Sorge's Law, the density profile r(z, t) will fluctuate with time t." It is to be expected that this variation causes differences mainly in the upper ∼10 m of the firn pack, i.e. the active layer, which is unsurprisingly where the largest differences between the Herron&Langway model and the ELM simulations is found (Figure 2 &3). Arthern2010 further notes that "Alternative models, broadly based upon the Herron and Langway [1980] parameterization, have employed different formulations for the sensitivity to temperature [Li and Zwally, 2004; Helsen et al., 2008]". This quote is a hint that the temperature dependence of Herron & Langway is not to be taken as the truth, which is kind of what the authors seem to be doing in Section 4.1.2, but also in the next section (4.1.2) where they calibrate model coefficients using their synthetic HL density profiles. I feel the heavy reliance on the HL model and synthetic data isn't properly justified.*
   Regrettably, we failed to provide enough details regarding the statistical modeling. Also, it is apparent that our evaluation method, i.e., direct comparisons of ELM output to the empirical model of Herron and Langway (1980), does not provide a convincing assessment of improved model performance. Unfortunately, we are not able to force ELM in single column mode like in CLM.
   To remedy these shortcomings, first, we are expanding the description of our statistical methods (i.e., Section 3.2) used to calibrate model coefficients. Second we are subordinating our reliance on the direct comparisons to the model of Herron and Langway (1980) used to justify ELM improvements. Section 4 now includes a geographical analysis from which we calculate root mean squared errors (RMSE) with reference to the extensive SUMup (measurement) dataset described by Montgomery et al. (2018). Based on this analysis, we find that applying our (imperfect but useful) statistical calibration to the current CLM snowpack and firn densification model reduces RMSEs for the majority of model grid-cells covering the Greenland Ice Sheet (GrIS).

2. *My second concern is with the readability of the manuscript, and in particular the range of different model configurations that are presented, and the purpose for all of them. The*

*title of the paper is "Snowpack and firn densification in the E3SM", however at the end of the manuscript I'm lost to which results are now representative of the improved E3SM and which aren't. Line 184 seems to suggest that the coefficients in the new E3SM are optimised from one of the other models, however it isn't stated which one, and the final configuration in E3SM is not named. The wording in Line 320 ("might expect") and Line 376 ("a first step") is also contradictory to this, suggesting that the coefficient optimization isn't really applied at all in E3SM. Does that mean that the entire Section 4.1.2 is actually superfluous and could be removed? Are none of the models discussed in this paper actually adopted in E3SM? I encourage the authors to make this more explicit. Alternatively, the authors could choose to take the focus off E3SM and shape it into a more general firn-modelling paper, I'll leave that up to them.*

This concern highlights an issue with our presentation of experiments and the lack of clarity regarding the final E3SM implementation. As such, the reader becomes confused by the end and can be left with several important questions in addition to those posed above.

To elucidate the highly experimental nature of our study, we are revising Section 3 so that it will enumerate ELM cases that each have a specific firn model configuration. Section 4 then presents results of these cases in the same order that they were introduced in Section 3. We are also splitting Section 4 into separate "Results" and "Discussion" sections and are restricting our more speculative remarks regarding future model development.

3. *My third major concern is with the comparison to observed firn core data (Section 4.2). Here the authors aggregate ELM data from across the GrIS from a coarse resolution simulation, and compare this a single point measurement (or at least, an approximation to this). Again, this seems to me like comparing apples to pears and a pretty crude approach. For instance, there will most certainly be grid cells in the composite that experience melt, whereas the interest is on dry firn compaction. Can the ELM not be forced in single-column mode with high-resolution meteorological data from e.g. ERA5, or a high resolution run with E3SM, more approximate to the actual weather at the site?*

   We originally relied on this observational comparison to correct obvious model biases because the measurements from Mosley-Thompson et al. (2001), which represent about 30 cores from various locations across the GrIS, exhibit much less variability than ELM density profiles throughout the same (approximate) domain of interest. Not only is this a crude approach, it also fails to provide a robust quantitative assessment of model realism that is missing in our justification of model improvement. Unfortunately, we are unable to force the ELM in stand-alone mode with higher resolution meteorological data. Given the current limitations, e.g., the low resolution ($\sim 2°$) forcing dataset and our one-dimensional snowpack model, however, our focus here is to reduce clear biases persisting globally in ELM when forced with plausible atmospheric reanalysis data. We believe that our experiments, while having limited capabilities, do show clear modeling errors that we attempt to correct with offline statistical modeling and validate with observations. In these comparisons, we exclude columns that undergo substantial melt by filtering out ELM grid-cells that have mean annual temperatures greater than -25 °C., as stated in the caption of Fig. 6.

   Again, to improve our evaluation methodology, we are expanding the results (Section 4) to emphasize a new geographical analysis that compares hundreds of measurements from the SUMup dataset, described by Montgomery et al. (2018), to the nearest corresponding ELM node in both space (via a location similarity matrix) and in time. This analysis stops short of a perfect "apples to apples" comparison, however, it does go further than our previous evaluation.

*All in all, I find this paper not convincing in its methodology, and I feel further justification or experiments are needed.*

We appreciate the constructive criticism and are working to address your valid concerns. We think you will find our expanded analysis more convincing and hope to provide more clarity in our forthcoming revision.

- *L61: pressure not pressures*
  We will correct.
- *L130: Actually, Muntjewerf et al. (2020) provides little detail on their model setup, and none on snow/firn modelling. Suggest to remove, and optionally replace with the following reference, which does actually provide more detail on snow modelling within CESM:*
  *van Kampenhout, L., Lenaerts, J. T. M., Lipscomb, W. H., Lhermitte, S., Noël, B., Vizcaíno, M., et al. (2020). Present-Day Greenland Ice Sheet Climate and Surface Mass Balance in CESM2. Journal of Geophysical Research: Earth Surface, 125(2), e2019JF005318. https://doi.org/10.1029/2019JF005318*
  We will update the reference accordingly.
- *L143: This title suggests that the new surface density scheme is only applied over ice sheets. However, this is not explained anywhere, so the title should be changed?*
  Perhaps you are correct. This sub-subsection title will be removed in our revision.
- *L180: Could you comment on what basis the values for T and A were selected? In Line 220 you define the "dry snow zone" as 0.5 m SWE / year, whereas here the value of 0.4 appears.*
  We selected the values based on Herron and Langway (1980). We reviewed their results to best determine where the model could be applied (in terms of mean annual temperature and accumulation rate) and focused our comparisons there.
  It is apparent that our basis needs to be better explained, probably in Section 3. In our revision, we are expanding the description of the statistical model, which relies heavily on the model of Herron and Langway (1980). Because we are subordinating the ELM experiment results to Herron and Langway (1980) comparison, we will add our basis for selecting temperature and accumulation domains in the expanded statistical modeling subsection. To add clarity, we will also comment on this domain selection in the discussion of the ELM experiment to Herron and Lanway (1980) comparisons.
- *L195: Could you comment on the quality of the meteorological data in CRU-NCEP over the regions of interest, i.e. ice sheets? Do you think the outcome of the simulations depends a great deal on the choice of meteorological forcing ?*
  The low resolution ($\sim 2°$) meteorological forcing provided by the CRUNCEP dataset is a primary limitation of our study. Because outcome of our ELM stand-alone simulations does, in fact, depend greatly on the surface boundary condition provided by the atmospheric forcing data, a mere geographical comparison of ELM to observations (which we are emphasizing in our revision) is not sufficient to disentangle density errors that are inherent from poor bound-

ary conditions versus those resulting from bad firn densification parameterizations. This is why we originally fall back on a comparison to Herron and Langway (1980) for evaluation, i.e., because it controls for direct drivers of the fundamental processes (accumulation rate and temperature) that determine advective strain rates and resulting density profiles.
By emphasizing a geographical comparison to observations in the revised "Results" section (4) and by moving our comparison to Herron and Langway (1980) to a new "Discussion" section (5), our revision offers two perspectives. First, we provide a direct look at how ELM density profiles compare to the real world and evaluate ELM realism quantitatively by computing RMSEs with reference to the SUMup dataset (Koenig and Montgomery, 2019). Second, we also dig into how well the firn model configurations are processing whatever boundary conditions they inherent from the available atmospheric input, independent of how well such conditions are reproducing the local climate. The latter is particularly important to evaluate in preparation for future climate simulations during which a warming climate will result in new, drifting boundary conditions for a given ELM grid-cell. Such conditions at a fixed location are not represented in observations, large-scale reanalyses, or historical reconstructions.

- *L196: please specify what nominal resolution (in degrees or km) does the ne11 resolution correspond to?*
  The horizontal resolution of the ne11 grid is 2.8°.
  As requested, we will provide the nominal resolution when referring to ne11.

- *L215: Since this manuscript concerns a global model, and there are only 16 layers to begin with, two reference firn density profiles are probably justified. Just be aware that there are more firn cores out there, e.g. see Figure 1 in the recently published TC paper by Verjans et al. :*
  *Verjans, V., Leeson, A. A., Nemeth, C., Stevens, C. M., Kuipers Munneke, P., Noël, B., & van Wessem, J. M. (2020). Bayesian calibration of firn densification models. The Cryosphere, 14(9), 3017–3032. https://doi.org/10.5194/tc-14-3017-2020*
  Thank you for the positive remark and the following referral.
  We are actually expanding the comparison of ELM experiments to measurements (including performance metrics) using the SUMup dataset (Koenig and Montgomery, 2019). We elaborate on this new analysis in our responses above.

- *L260: remove comma after ELM ?*
  Removed.

- *L 373: near-surface firn densities that are too low, not large?*
  In A'76 (ELM v1), surface densities are too low. Just below the surface, densities are too high. This finding supports results from van Kampenhout et al. (2017) discovered in CLM version 4.
  We understand that the distinction here between "near-surface" and (e.g.) just "surface" is unclear. We will rephrase this and similar sentences throughout our revision to state more specifically where along the depth coordinate our findings apply.

- *L 376: E3SM Project: not clear, is this a reference?*
  Yes, this is a reference. We will change this to just "...within E3SM."

- *L 380: and AIS? Surface melt is believed (or known) to be important for the stability of ice*

*shelves.*
Ice shelves are not within the ELM domain and thus are not the focus of our present study.

- *Reference list: to avoid cluttering, I'd suggest to remove the URLs and replace with DOI where needed.*
  Regarding the reference list, we will follow the guidelines and whatever recommendations given by the editorial support staff.

- *L 446 : fix title L 483 : fix title L 512 : fix title*
  Thank you for bringing these fixes to our attention.
  We will review (and fix) these titles in our revised manuscript.

- *Figure 4: The caption describes the meaning of the line graphs in the first row, but not the crosses used in the second row.*
  We are revising Fig. 4.
  We will also update the caption appropriately and add a description for the crosses.

- *Figure 5: The titles of the subfigures could be made more informative. Also, the colour bar could be made more restrictive, it appears densities $< 300$ kg/m3 do not occur at all.*
  We will provide more informative titles and consider changing the color bar.

- *Figure 6: The legend appears not consistent with the previous figures, e.g. ELM-A'10 instead of just A'10.*
  Our revised Sections 3 and 4 and our new "Discussion" Section (5) will be consistent in the labeling of ELM experiments.
  We will also update all the figure legends accordingly.

For a direct look at how we address these concerns and comments, please see the forthcoming revised manuscript.

Sincerely,

Adam M. Schneider et al.