Review of Medley et al. by Vincent Verjans.

The study of Medley et al. estimates Surface Mass Balance (SMB) and Firn Air Content (FAC) evolution over the time period 1980-2021, on both the Greenland and Antarctic ice sheets (GrIS and AIS). This study represents a comprehensive modeling effort to derive model estimates of elevation changes associated to surface processes. The authors have thoroughly reworked the manuscript since its first version. This includes revision of some modeling aspects (use of effective temperature, degree day modeling, surface density parameterization), uncertainty analysis, extensive comparison of SMB against state-of-the-art studies, and improved discussion of limitations. I believe that the authors have appropriately addressed my comments from the first round of revisions, as well as those of reviewer 2. This review is separated in minor comments and technical comments. My minor comments address some aspects that should be clarified, or about which I raise some secondary reservations with respect to the approach and/or judgment of the authors. My technical comments only address structural and language aspects of the manuscript. I sincerely appreciate the thorough work of the authors to address all the comments from the first round of reviews. Provided some minor issues in the updated manuscript are addressed by the authors, I encourage the publication of this study in The Cryosphere.

Minor comments

1) The interpretation of the Reference Climate Interval (RCI)

1.a)

The authors define the RCI as "*ideally representative of long-term steady-state conditions*". I believe that this interpretation is slightly wrong. The RCI is used to develop the initial model firn column, from which transient experiments over the period of interest (1980-2021) start. As such, ideally, the initialization should be computed with the true climate forcing of the decades and centuries preceding 1980. This is true regardless of whether the long-term conditions were in steady-state (i.e., without trends) or not. In other words, the "perfect" RCI should not represent steady-state conditions if the true conditions were not in steady-state prior to 1980.

As the authors rightly point out:

"*we only have a spatiotemporally complete understanding of polar climate conditions arguably since the beginning of the satellite era (1979 and onwards). Thus, we make assumptions regarding how that firn column will respond to modern conditions (…)*"

In light of this incomplete knowledge, using steady-state conditions over the RCI is only the most reasonable assumption possible to make, but not a necessary condition for a valid firn model initialization procedure. For this reason, I believe that the authors should revisit their discussion of the RCI and of the appropriateness of their assumptions in Sections 2.1.2, 3.2.1, and 3.2.2. It should be clearer that steady-state conditions over the RCI are used in order to isolate effects of climatic deviations from the RCI on firn column changes. But such steady-state conditions are not representative of true conditions, and the true changes in firn thickness are influenced by the unsteady nature of past climate conditions. This is already partly explained in the Discussion section, but I find that earlier statements in the manuscript are misleading. Furthermore, the Discussion section only discusses past trends on AIS, whereas the existence of past trends is also true for GrIS (e.g., Hanna et al., 2011), which should also be mentioned.

1.b)

In the Discussion section:

"*we expect our results as a lower bound for trends, and future work investigating the impact of these reconstructed trends would help to quantify the resulting uncertainty in height changes due to long-term climate change*"

Either I misunderstand this statement, or I respectfully disagree with the authors about the lower bound for the trends. The presence of spatial trends in past climate does not mean that assuming a steady-state RCI necessarily causes an underestimation of recent trends in firn thickness change. That is because recent trends are relative changes compared to firn column dynamics during the RCI. For example, if the RCI overestimates SMB in past climate with respect to reality, a decrease in SMB in the recent past

would cause a stronger surface lowering than what happened in reality. That is because the relative decrease in SMB in the simulations would be larger than the relative decrease in SMB in reality. Thus, computed trends are not necessarily a lower bound. Instead, an RCI not representative of the true past climate can result in both an over- or under-estimation of current trends, depending on the particular biases of the RCI assumptions (which vary in space, and are different for different variables).

2) The uncertainty analysis

In general, I appreciate the effort to perform the uncertainty analysis in this revised version of the manuscript. The method is robust, and the uncertainty values add a lot of value to the Figures 13, 16, and 18. I would like however to bring three minor points to the attention of the authors.

2.a)

It is not entirely clear to me how the authors sampled the values of the variables in their uncertainty analysis. They state that:

"*we sampled the 2-sigma Gaussian distribution error in the modelled initial density (rho_0) and the calibration parameters (alpha_0, alpha_1, E_c0, E_c1) and perturb the CFM parameters*" (Section 2.5.1)

"*The perturbation developed sample randomly within the 2-sigma bounds of the Gaussian Perturbations and from a small number of Random Perturbations.*" (Table 1)

Does that mean that (i) they sample from a Gaussian distribution or (ii) from a Uniform distribution bounded by the 2-sigma bounds of the uncertainty?

I guess that they did (i), but in that case I do not understand why they limit the sampling to the 2-sigma bounds and thus exclude outliers from the uncertainty analysis. Alternatively, I may be misunderstanding the method. I recommend that the authors clarify and justify the method.

2.b)

Uncertainty in the parameters, and even more in the climatic variables, are certainly correlated. For example, SMB perturbations are, in reality, negatively correlated with perturbations in temperature. Similarly, the model parameters are not independent of each other. I understand that the authors decided not to constrain the appropriate correlation values, and their approach is sufficient in my view. However, I ask the authors to mention that there is some dependence between the perturbed variables, and that this is not accounted for in the uncertainty analysis.

2.c)

In Figure 12, there are 55 points plotted, whereas Section 2.5.1 mentions 45 sites for AIS and 18 sites for GrIS (i.e., total of 63 sites). Why is there a discrepancy of 8 sites?

Technical comments

In general, there are a lot of places where commas should be added for better clarity. I identify some of them in my Technical comments, but it would be good to pay attention to missing commas when the authors re-read the manuscript.

p.1 l.19-20

Change "*associated with surface mass balance*" to "associated with mass fluxes from surface processes". I want the readers to keep in mind that firn air content fluctuations themselves are largely governed by surface mass balance fluctuations.

p.2 l.44

Add "constant": "becomes approximately constant (917 kg m-3)".

p.2 l.62

Add comma: "changes, yet".

p.4 Eq. (6)

Remove brackets from the numerator.

p.4 l.111

Typo: "*GSDC-FDMv1.2*".

p.5 l.135-136

"*This depth is divided by a burial rate (snowfall – sublimation – melt) to estimate the time needed to refresh the firn column for a given site*."
I believe that "*This depth*" should be replaced by "The cumulative mass until this depth".
p.6 l.157
Remove "*that*".
p.6 l.161
I suggest replacing "*total mass above*" by "cumulative accumulation above" because b does not account for mass removal via runoff.
p.6 l.169
Add "is": "and is based".
p.6 l.179
Add "the": "on the use".
p.7 l.183-185
The iterative 3-sigma edit method is not clear to me. "*removing individual density measurements with residuals to the linear model larger than 3-sigma*": Larger than 3-sigma of what? Is "*sigma*" here the root mean square deviation of all the individual density measurements with respect to the linear model? In this case, it should be clarified that "*sigma*" refers to the root mean square deviation.
p.7 l.195
Add a space: "(2010) model".
p.8 l.232
Add commas: "parameters, when plugged into Eqns. 11–12, provide".
p.8 l.235
Add comma: "will increase, while".
p.8 l.242:
Replace "*expectation*" by "assumption".
p.8 l.244
Add comma: "with increasing depth, T".
p.9 l.246:
Replace "*expectation*" by "assumption".
p.9 l.255-256
I am not sure that this sentence is correctly phrased.
p.10 l.280
Add commas: "that, on average, we".
p.10 l.281
Plural: "deviations".
p.10 l.289-290
Give RMSE of fit of rho_0 for all rho_0 values as well as for rho_0 values below 330 kg/m^3.
p.11 l.311
"*GEOS*" is not defined.
p.11 l.315
Change "*span*" to "spans".
p.12 l.347
"*normalized distance*": normalized to what? To the mean r^2 and RMSE of the grid cell?
p.12 l.350
Add "an": "an observation-based calibration data set".
p.12 l.360
Add "in": "more in Sect. 4.".
p.12 l.365-366
Add a comma: "0.13 kg m^-2 hr^-1 K^-1, while calibrated values".
p.13 l.369
Change "*complicates*" to "complicate".

p.13 l.372
Why set the DDFs lower than the lower bound of ice shelves to 0 and not to the lower bound itself?
p.13 l.374
Provide also the range of DDFs in Antarctica. And maybe give a brief comparison with the Greenland values.
p.13 l.382-383
Rephrasing needed.
p.13 l.385
Change "*reduced the mismatch*" to "caused a larger mismatch". Or clarify the sentence.
p.13 l.388-391
Yes, but mention that the calibrated firn model is nevertheless used in areas with meltwater percolation.
p.13 l.397
Add space: "For v1 and".
p.14 l.406
Change "*error analysis*" to "uncertainty analysis".
p.14 l.420
I recommend replacing "*densification rates are reduced under increasingly high accumulations*" by "the sensitivity of densification rates to increasing accumulation is reduced".
p.14 l.420
Change "*dramatic*" to "pronounced".
p.14 l.421
Add comma: "is increased, especially".
p.15 l.441:
Change "*or*" to "i.e.,".
p.15 l.446
Add "of variables": "time series of variables of critical importance".
p.15 l.454
Add "the": "Each of the perturbations".
p.15 l.455
Add a comma: "Gaussian distribution, except".
p.15 l.458-459
Snow accumulation should be defined as Sn-Ev for consistency with Eq. (5).
p.16 l.483
Make sure to use the same tense for the verbs.
p.16 l.495
Refer to Section 2.5.2 when introducing the observations of SMB.
p. 17 l.500
Add hyphen: "SMB-induced".
p.17 l.516-524
How did the number of observations reduce from 16427 to 1037? Is that because many observations fall within the same grid cell? And/or are there observations excluded because they do not span a long enough time period? Or for another reason?
p.17 l.530
Rephrasing needed.
p.17 l.531
What is meant by "*accumulated*"?
p.18 l.533
"*First, we determined the mean GSFC SMB over the exact observation interval.*"
This comparison is never analyzed or discussed in the remainder of the manuscript. I recommend removing it for the sake of clarity.
p.18 l.552

Add “on average”: “the AIS firn column contains, on average, more air than the GrIS.”.
p.18 l.560
Specify that runoff is an output of the CFM, and not of MERRA-2.
p.19 l.564
“*the firn column accommodated 40% of all liquid water*”: give +/- annual variability of the percentage.
p.19 l.569
“*The RCI is ideally representative of long-term steady-state conditions*”: see my Minor comment 1 about the “*steady-state*” aspect.
p.19 l.574
Change “*most likely*” to “significantly”.
p.19 l.575
Change “*our choice of RCI (1980–1995) should not generate non-physical transients in our firn simulations*” to “our choice of RCI (1980–1995) should not generate transients associated with the initialization process in our simulations.”
p.19 l.579
“*The firn column only accommodated 38% of liquid water*”: give +/- annual variability of the percentage. Also, please specify if the decrease from 40% to 38% is statistically significant.
p.19 l.580
“*The ablation zone grew in area by 30%*”: please specify how this was computed. Does this number come from a fitted trend on the area with SMB<0 over the annual time series?
p.19 l.588
“t*he firn only accommodated ~19% of all liquid water*”: give +/- annual variability of the percentage.
p.20 l.594
“*majority (94%)*”: give +/- annual variability of the percentage.
p.20 l.601-602
Change “*the choice of RCI is justified*” to “the firn column initialized over the RCI spin-up should be in equilibrium with steady climate conditions”.
p.20 l.607
Typo: “*and*” should be “an”.
p.20 l.613
Change “*cycles*” to “amplitudes”.
p.20 l.619
Change “*skewed*” to “driven”.
p.21 l.623
“*(86.3 ± 13.6 km^3 yr^-1 )*”: what do these value refer to?
p. 21 l.626
Change “*component*” to “amplitude”.
p. 21 l.646
Add “mean”: “than the ensemble mean”.
p.21 l.647
Add “observations with”: “for observations with SMB > ~ -2”.
p.22 l.654-655
Also give the value of the mean absolute relative bias for the sake of information.
p.22 l.659
Add “integrated”: “suggests that integrated over the entire ice sheet”.
p.22 l.675
Typo: “*GSFC the*”.
p.23 l.688-690
Provide the results of the ensemble mean of Mottram et al. (2021) in Table 3. Otherwise, the reader cannot evaluate the comparison in a quantitative manner without going to the publication that is referenced.

p.23 l.695-697
Also give the value of the mean absolute relative bias for the sake of information.
p.23 l.704
"*exceeded by two models within the ensemble*": two of how many?
p.23 l.705
Change "*within*" to "in".
p.23 l.708
Add "compare": "We also compare".
p.24. l.725
Change "*largely in response to the overburden*" to "largely due to reduced sensitivity to increasing overburden".
p.24 l.727-728
"*Our calibration differs*": note that Verjans et al. (2020) took a similar approach.
p.24 l.742
Add comma: "firn densification model, which models".
p.24 l.766
Change "*recent past*" to "past prevailing conditions".
p.25 l.775
Add hyphen: "physically-based".
p.25 l.776-778
I believe that the most likely cause of the lower melt values is that the calibrated DDFs are capped at higher elevations after the calibration process. This should also be mentioned here.
p.25 l.782
Typo: "*heigh-elevation pats*".
p.25 l.782
Add comma: ", particularly".
p.26 l.785-786
Add an extra reservation by changing: "*to fully evaluate this improvement and highlight other potential future improvements.*" to "to fully evaluate this improvement, rule out possible compensating errors, and highlight other potential future improvements.".
p.26 l.800
Change "*e.g., fresh snowfall*" to "i.e., fresh snowfall".
p.26 l.802-803
"*we want to separate the climate model impact (SMB) from the firn model impact (FAC)*": I respectfully disagree with this statement. Firstly, air changes due to snowfall are governed by the climate model, not by the firn model. Secondly, the firn model takes as inputs the fields from the climate model, thus the effects from both models cannot be separated entirely.
p.26 l.803
"*While the SMB and FAC contributions to total firn volume change over multiannual time scales are somewhat comparable*,": I believe that this does not give an appropriate picture of the results given in Sections 3.3.1 and 3.3.2. I think that it would be better to add one or two extra sentences to quantify the different impacts of SMB and FAC on long-term volume changes on the GrIS and AIS.
p.27 l.824-825
Typo: "*within firn column*".
p.27 l.826
Add "in-situ": "measuring firn processes in-situ".
p.33 l.838-839
I believe that the authors calculate a series of 141 RMSE values for stage 1 and 76 RMSE values for stage 2, from which they calculate the lower quartile, median, and upper quartile RMSE. Similarly for r^2. Is that correct? If so, I recommend adding a statement such as: "For each observation, we use all the point measurements of density in depth to calculate the corresponding RMSE and r^2."

p.33 l.839:
Typo: "*observation*" should be plural.
p.33 Table B1
Change the column "*Density Profile*" to "Linear Density Profile".
p.34 l.847-848
"*using the daily MERRA-2 fields*": this contradicts the statement above "*native 1-hour resolution*". Are the fields from MERRA-2 hourly or daily?
p.34 l.848
Change "*label*" to "labeled".
p.34 l.854
Specify that n is taken equal to 5.
Figure 1
Explain the difference between red and black circles. Also, most red crosses are not clearly visible.
Figure 2
At first, it was not intuitive to me why the contour plot has the same color code as the closed circles. I recommend adding a statement such as: "The background contours represent the best fit to the coefficients from the calibration sites.".
Figure 4
"*those in grey reference the open circles*": I don't think that "reference" can be used as a verb.
Figure 5
I recommend plotting these figures as differences between MERRA-2 and M2R12K instead.
Figure 6
I recommend plotting these figures as differences between MERRA-2 and M2R12K instead.
Figure 7
"*as it maximizes the distance between the two curve*": that depends on the y-axis scales that are chosen for the respective curves. Instead, this should be reformulated as minimizing the normalized distance (and see my comment for p.12 l.347). Note also that "*curve*" should be plural.
Figure 11
Is it possible to show these plots as box plots? The information conveyed would be similar, but that would also allow the reader to see the outliers.
Table 1
Rephrasing of the caption is needed. And reference to Calonne (2019) should be to Calonne et al. (2019), and it is not given in the references.
Table 3 (l.1389)
Typo: "*difference*" should be "different".

<u>References used in this review:</u>
Hanna, E., Huybrechts, P., Cappelen, J., Steffen, K., Bales, R. C., Burgess, E., McConnell, J. R., Peder Steffensen, J., Van den Broeke, M., Wake, L., Bigg, G., Griffiths, M., and Savas, D.: Greenland Ice Sheet surface mass balance 1870 to 2010 based on Twentieth Century Reanalysis, and links with global climate forcing, J. Geophys. Res.- Atmos., 116, D24121, doi:10.1029/2011JD016387, 2011.

Verjans, V., Leeson, A. A., Nemeth, C., Stevens, C. M., Kuipers Munneke, P., Noël, B., and van Wessem, J. M.: Bayesian calibration of firn densification models, The Cryosphere, 14, 3017–3032, https://doi.org/10.5194/tc-14- 3017-2020, 2020.