Summary of study and overall assessment

In this study, the firn properties of Greenland are simulated for the period 1980 – 2020 with two different models – the semi-empirical Community Firn Model (CFM) and the physically-based SNOWPACK (SP) model. Both models are driven with atmospheric forcing from MERRA-2 reanalysis and applied on the same spatial grid (~0.5°) as the reanalysis data. To equilibrate the firn layer, the models were first spun-up with forcing data from the so-called reference climate interval (RCI), which ranges from 1980 to 1995. Subsequently, the actual simulations with CFM and SP were performed and evaluated with 767 firn cores from the SUMup project. After demonstrating the good overall performance of both models, results are analysed and intercompared with a focus on firn air content (FAC) and its temporal evolution (interval means, inter-/intraannual changes) and spatial distribution on basin scales.

Firn models are important tools, because they allow a spatially comprehensive assessment of ice sheet's average firn porosity. This quantity is essential for estimating the potential of the ice sheet to retain meltwater in the firn layer (buffer effect) and thus slowing down global sea level rise. It is therefore crucial to have a good understanding of how well firn models of different complexity simulate firn related processes. This manuscript adds interesting results and findings to previous work. The study is generally well written and structured and the figures are of excellent quality. Find below some suggestions to improve the manuscript – most comments are of minor nature and concern details.

We thank the reviewer for taking time to read and comment on this manuscript. We are especially grateful for the thoughtful and inspired questions about this work. The updated manuscript has been greatly improved following the reviewer's comments. We would also like to note that we have changed "CFM" to "CFM-GSFC" per another reviewer's suggestion, so the responses below contain the latter use. The reviewer's comments are in black text, the authors' responses are in blue text, original manuscript text that has been removed or modified is *"blue and in quotes and italics"*, and new manuscript text is ***blue and in italics and bold***.

General comments

Conclusion section

In my opinion, the conclusion section needs some improvement. The structure seems currently a bit chaotic – e.g. the part with the outlook ("This will in turn allow us to better predict the firn's response to future warming.") should rather be at the end of the section. I suggest to rearrange this section in a more logical way.

We thank the reviewer for this suggestion, and we have restructured and refocused the conclusion section. Below we address some of the specific points noted and show how we have updated the manuscript.

Furthermore, the following points could be included/extended:

- Embed findings in a larger picture (and discuss further implications). For instance, I guess the computational cost of running the physically-based SNOWPACK model is substantially higher (could you state how much approximately?). Does the higher complexity (e.g. explicit consideration of effects like wind compaction under drifting/blowing snow that influence new-snow density) "pay off" (i.e. add some distinctive benefits)?

Thank you for this suggestion. The computational costs are difficult to evaluate due to several factors including the fact that the models are being run on different systems, changing parameters in either model can drastically affect runtime, and different climate conditions (e.g., high melt) can affect model runtimes differently. As such, neither model necessarily stands out as being more computationally efficient or inexpensive. The SNOWPACK and CFM-GSFC simulations performed for the current work took about the same amount of time to perform. However, as to your second point, we have summarized some of the potential benefits of each model in the conclusion section. Our results do not support the idea that one model is better than the other, but rather quantify their differences in several contexts (e.g., stable climate, changing climate, melt, etc.). To the conclusion, we have added:

In the present work, neither model clearly outperforms the other within the scope of the evaluation. Observational limitations prevent the recommendation of one model over the other,

but here we report potential benefits and drawbacks of each model. The physics-based design of SNOWPACK means that it is not tuned to observations and consequently not biased toward available observational data. This may mean that it can more realistically simulate firn properties under future climate conditions whose effects are not capture in existing firn observations. However, firn physics are not fully understood and knowledge gaps limit the accuracy of the model. The CFM's modular design allows for the user to easily choose from several densification schemes. Semi-empirical densification schemes such as the one used in the present work are tuned to observations, which means that realistic densification relationships are built into the model and there is less need to rely on poorly understood physics. To recommend one model over another, additional evaluation data and metrics would be necessary.

- State recommendations for future (similar) studies and extend outlook. For instance, which are the most crucial processes in firn model that should be better represented in future models (I have in mind processes like vertical (or even lateral) water flow, reduced permeability of ice layers/slabs, ponding water conditions in firn aquifers, etc.)?

These are critically important ideas and questions for the firn modeling and observation communities. Capturing these complex meltwater processes in observations is difficult, and the physics are also not well-constrained. This makes it difficult to determine which process is most crucial to model accuracy. It also makes it difficult to accurately represent these processes in models if the real-world examples are limited. As such, we have added the following to the conclusion section in order to share what may be needed in order to more accurately capture certain processes in firn models:

Moreover, firn models in general are limited by knowledge gaps in firn hydrological processes such as vertical meltwater percolation, lateral flow, and conditions for firn aquifer and ice slab formation. Additional research focusing on obtaining detailed observations of these processes would provide opportunities for important developments in firn modeling.

Point-comments

Content-related (text)

Line 10: For which time are these statements valid? 1980, 2020 or averaged over the 40 years?
Thank you for noting this. We have added ***1980–1995 average*** to clarify.

L84: I would call this section "**Methods and data**" (because you also present the SUMup observations)
Done.

L88: Do you consider both snow- and rainfall data from MERRA-2? Or do you derive precipitation fractions (solid/liquid) with an air temperature threshold?
From MERRA-2, we obtain the precipitation as three variables from the integrated diagnostics of water and energy dataset (GMAO, 2015). These are the convective rainfall, large-scale rainfall, and snowfall. To make this clearer to the reader, we have added ***(rainfall and snowfall)*** following *"precipitation"* in the list of MERRA-2 variables used.

L95: I would explicitly state that MERRA-2 was also considered in Zhang et al. (2021) – this is not obvious from the current statement. Maybe you could also briefly summarise how the model performs with respect to Automatic Weather Stations (AWSs) data.
Thank you for this suggestion. Per another reviewer's suggestion we have removed this sentence and replaced it with a broader and less biased one: ***A different reanalysis product or regional climate model could also be used here, though the choice of forcing dataset will not affect the firn model intercomparison since we provide the two firn models with identical input.***

L99: I have a general question (just out of curiosity – no changes regarding this question are required for the current manuscript): SNOWPACK and CFM inherit MERRA-2's spatial grid. However, one could also apply a different (unstructured) grid, which e.g. has a higher spacing close to the ice sheet's margins. With this, one could better capture areas with strong climate gradients and the complex boundary of the ice sheet (which might also reduce the disagreement in total glaciated area). However,

such a solution might anyway only be relevant if a generally higher grid spacing than 0.5° is used (also in terms of atmospheric forcing data). What's your option on this idea for future firn model applications to the GrIS?
We thank the reviewer for this insightful question that could inspire a future research project. A finer resolution grid would indeed be a way to capture the fine spatial-scale impacts of climate and topography on firn structure, especially in these marginal areas. The limitation here would be the accuracy of the climate model. The marginal areas are notoriously difficult to represent as they are characterized by steep topography, orographic precipitation, strong temperature gradients, complex winds, and flowing outlet glaciers. As such, the choice of an adequately accurate downscaled or finer scale climate model would be difficult. Still, this is an interesting question that perhaps could be investigated by using a suite of regional climate models and focusing the modeling/analysis on the margins.

Section 2.2: Could you specify which scheme for vertical water percolation is applied in SNOWPACK?
Yes; thank you for pointing out that we have forgotten to mention SNOWPACK's percolation scheme. We use the bucket scheme for both models, and we have updated the text to include: ***We apply a bucket scheme to represent vertical water percolation in SNOWPACK.***

L125: It might be useful to refer to Fig. A2 here (time series in the grey-shaded areas show no (strong) temporal trends, which supports the definition of the RCI period)
Thank you for this suggestion. We have added ***(Fig. A2)*** to this line.

L135: Why do you perform the vertical interpolation only for CFM output (and not for SNOWPACK – which also has a fine grid spacing)?
Here, the different choices in whether or not to vertically interpolate comes down to differences in the way we chose to reduce computational costs. Both models actually have a layer-merging scheme that reduces computational costs. In the CFM-GSFC, we reduced file sizes by interpolating onto a regular grid, and in SNOWPACK, we only save output every 7 days. These were choices made early in the project, and we may reconsider these if we were to run the models again.

L150: Why do you apply different spin-up conditions for SNOWPACK and CFM? Is it due to computational constraints (i.e. that SNOWPACK is more expensive to run)?
To the best of our ability, we have made the spin-ups as close to identical as possible. Any differences in the spin-up regimes are just due to the way the models were coded and designed. We have removed *"For example, if the firn needs 1000 years to spin up, the RCI would repeat 63 times. Once the spin up is completed, the main model run (1980–2020) commences."* to reduce any confusion.

L169: Here, you neglect any liquid water in the firn – right? Compare e.g. to Eq. (6) and (7) in Kuipers Munneke et al. (2015).
Yes, that is correct. We calculate FAC as in Eq. (7) in Kuipers Munneke et al. (2015).

L174: Why do you use 100 m as a lower limit (and not e.g. 150 m – the spin-up depth of SNOWPACK)?
We choose 100 m as a cutoff for two reasons. (1) Most of the simulations reach solid ice by 100 m, meaning that there is negligible FAC below 100 m for many cases. (2) We chose a depth that every simulation in both SNOWPACK and the CFM would reach. In SNOWPACK, the spin-up will be complete when there is either 150 m of firn or when the bottom 3 m is solid ice (see Section 2.2). This means that a simulation may not reach 150 m, but we calculated the minimum thicknesses from all simulations (both models) and found that they all reached at least 100 m depth.

L181: I would briefly explain what the NSE range (<0, 1, etc.) means for the model (because most readers are probably unfamiliar with this metric)
We thank the reviewer for this suggestion. We have added ***A value of 1 indicates perfect model performance, whereas a value of 0 indicates that the model's predictive ability is the same as using the observations' means.***

L269: “no change” might be a bit too restrictive. Maybe better “only negligible changes”
Thank you for this suggestion. Per another reviewer’s suggestion and this reviewer’s comment on Line 323, we have chosen to replace Table 2 (and the explanation of its results) with trends during the two compared periods (see below). Thus, we have removed this line.

Table 2. Modeled spatially-integrated firn air content (FAC) trends and standard errors of the trends for each of the six basins (Fig. 6) for the 1980–1995 reference climate interval (RCI) and the 2005–2020 period. The last row shows the trends for the full GrIS.

	1980–1995		2005–2020	
	CFM-GSFC	SNOWPACK	CFM-GSFC	SNOWPACK
Basin	trend (km^3 y^{-1})	trend (km^3 y^{-1})	trend (km^3 y^{-1})	trend (km^3 y^{-1})
NW	-0.1±0.1	-1.1±0.1	-2.6±0.1	-12.3±0.2
CW	+0.2±0.0	+1.5±0.1	-2.1±0.0	-10.3±0.1
SW	-1.0±0.1	-1.2±0.2	-6.9±0.1	-16.7±0.2
NO	+0.6±0.0	+0.4±0.1	-5.7±0.0	-11.4±0.1
NE	-1.3±0.1	-0.8±0.3	+1.9±0.1	-2.5±0.2
SE	+3.1±0.1	+3.1±0.3	+2.4±0.2	-4.5±0.4
GrIS	+1.2±1.0	+1.7±0.9	-17.4±1.2	-66.6±1.2

L289: I would shift this first paragraph (maybe to the end of this section?). For me, these first lines suggest that it is not interesting to look at trends because there is no significant change in FAC between the two periods. However, looking e.g. at Fig. 8, there seems to be a clear trend during the latter period which is definitely worthwhile to discuss. Anyway, I have to admit that I’m not an expert on statistical methods, so there might be a reason why you start with comparing the two periods statistically…
We appreciate this comment and suggestion from the reviewer. We chose to report these results here for the sake of thoroughness and to create a baseline for comparing with the later 2005-2020 period, which as noted by the reviewer, is much more exciting. To make this clearer to the reader, we have changed the first line of this paragraph to: ***To examine how the models represent the seasonal cycle in spatially-integrated FAC during the RCI,...***

L323: I think it would be more robust to look at linear trends here. Computing the difference between two (somehow arbitrary selected years) is prone to noise introduced by interannual variability…
This is an excellent point and we see the limitations of comparing two single years, especially in relation to the interannual variability. As such, we have shifted the focus from comparing two years to comparing the trends between the two periods (1980–1995 and 2005–2020). We have remade Table 2 (shown above in an earlier response) and replaced this paragraph with:

The marginal areas of the GrIS experience the greatest amount of FAC depletion between 2005 and 2020 (Fig. 10). Both models simulate the same spatial patterns in loss, but the trends vary by basin (Table 2). SNOWPACK simulates a negative trend in spatially-integrated FAC in all basins during this time, with the strongest trend of -16.7±0.2 km^3y^{-1} in the southwest. The negative trend is weakest in the northeast (-2.5±0.2 km^3y^{-1}) and southeast (-4.5±0.4 km^3y^{-1}), which are also the only two basins where the CFM-GSFC simulates positive trends (1.9±0.1 and 2.4±0.2 km^3y^{-1}, respectively). The CFM-GSFC also simulates the strongest negative trend in the southwest where the spatially-integrated FAC change is -6.9±0.1 km^3y^{-1} (Table 2).

L445: I’m not able to follow this sentence. Do you mean “intensified firn densification”? And why does that increase the firn’s cold content?

Thank you for noting this point of confusion. We have rewritten this sentence for clarity, and also removed the language about cold content since it was tangential to the main point. It now reads: ***Pore space depletion can also be caused by firn densification, which in turn modifies the meltwater refreezing and retention capacities of the firn in a complex manner (Vandecrux et al., 2020).***

Typos, phrasing and stylistic comments
Line 6: …Community Firn Model (CFM)**,** to quantify…
Done.

L15: This sentence reads odd somehow. It might be better to add the negative rates to the previous sentence and then state: "The reduction in spatially-integrated FAC in SNOWPACK and CFM demonstrate how model differences propagate throughout the FAC record."
Thank you for this suggestion. We have changed these lines to: ***During this period, the spatially-integrated FAC across the entire GrIS decreases by 3.2 % (-66.6 km^3 y^{-1}) in SNOWPACK and 1.5 % (-17.4 km^3 y^{-1}) in the CFM-GSFC. These differing magnitudes demonstrate how model differences propagate throughout the FAC record.***

L117: "scheme use to" "scheme use**d** to"
Done.

L195: "form**ed**"
Done.

L197: "in the surface" "close to the surface"?
Changed to ***near the surface***. Thank you.

L364: I would change this to something like: "The five locations shown in Figure 4 lie all within the same MERRA-2 grid cell and thus share the same atmospheric forcing data for the models."
Done. Thank you for this sentence restructuring; it reads much better now.

L366: change "MERRA-2 grid point" to "MERRA-2 grid cell" (also later in the text)
Done. We have made 7 replacements.

L374: "in simulating observations" "in reproducing observations"
Done.

L375: I would rephrase this sentence.
This has been changed to: ***Compared to the dry and flat ice-sheet interior, areas with firn aquifers or steep topography are likely to have higher model uncertainty.***

L443: I would rephrase this sentence.
This has been changed to: ***The largest percent change is in the southwest, which has the warmest temperatures and highest melt compared to other basins during this period (Table A2).***

L461: I would rephrase this to something like: "For both models, the summer air temperature seems to be a good proxy for the abrupt drop in FAC, which happens at temperatures between approximately -4 to 0° Celsius."
We have changed this sentence to: ***Modeled FAC abruptly decreases at a summer air temperature threshold of 4°C, and the models remain in agreement in warmer conditions.***

Figures and Tables
Figure 1: Adding degree symbols and N/E to the latitude/longitude coordinates would help the reader.
Thank you for this suggestion. We have added degree symbols to the coordinates, but there is not much room in the legend for N/E, so we have kept with the positive/negative convention. Updated Figure 1 is below:

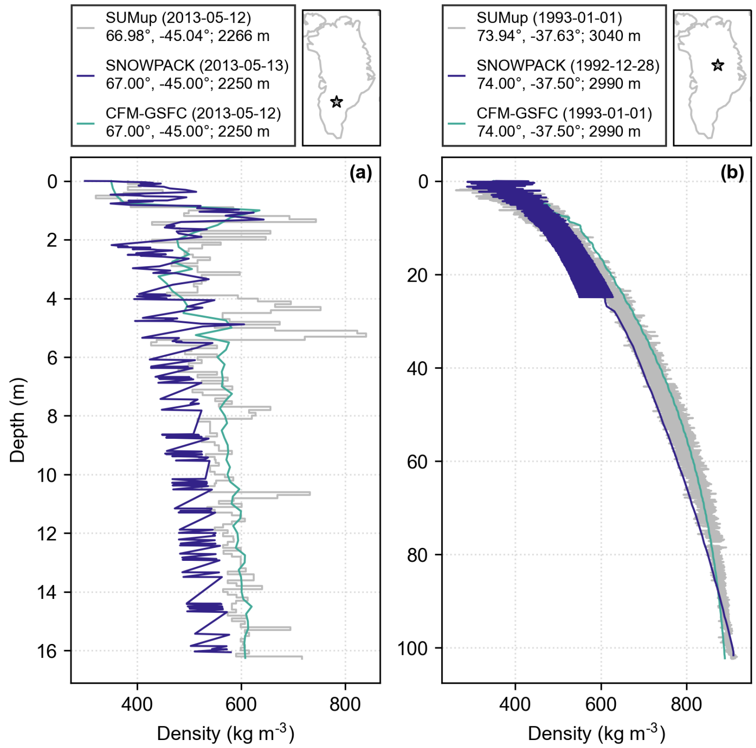

Figure 2: I would state relative biases in percentages like specified in Eq. (4)

Thank you for this suggestion. We have change the fractional biases to percentages to better reflect Eq. (4). The color scale has also been changed per another reviewer's comment. Please see the updated Figure 2 below.

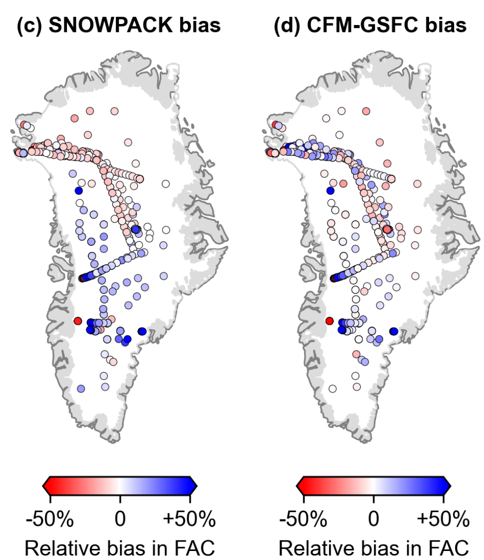

Figure 4: I would change "MERRA-2 domain" to "MERRA-2 grid cell" and in the caption: "MERRA-2 grid point" "MERRA-2 grid cell"

Done. The updated Figure 4 is shown below.

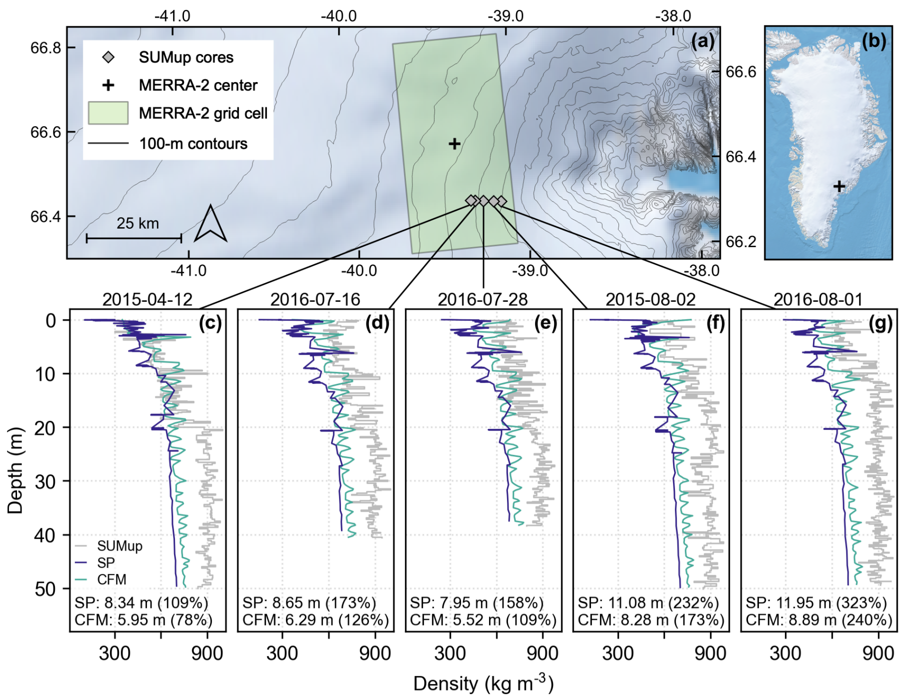

Figure 6: caption what caused the missing data?
We have added the following explanation in Section 3.2: ***A few areas of missing data exist and are due to one or both of the firn models encountering an error in the simulation (Fig. 6). For example, if a grid cell is located in the ablation zone and does not receive enough accumulation to model a snow layer, SNOWPACK will stop the simulation and report an error.***

Table 3: How did you distinguish between detectable and undetectable signals?
We have added the following text to Section 3.2 to clarify why this is: ***(i.e., some basins contain too much intra-annual variability for the sine fitting function to detect a seasonal cycle)***

Fig. A1: caption: this means you only consider SUMup observations for this analysis in which the upmost density measurement covers the topmost 0.1 m or less – right? Furthermore, I would change the following sentence slightly: “The CFM uses a prescribed surface density of 350 kg m^{-3} **(green vertical line)**, which falls near many of the observed surface densities.”
To the first point, yes, for this figure we only consider SUMup cores that have a measurement within 0.1 m of the surface. This is not the case for the entire paper’s analysis though, so we have added to the caption, ***in this figure*** to clarify. To the second point, thank you. We have restructured that sentence. The full, updated caption now reads:

Figure A1. (a) Observed surface density (ρ_0) from SUMup versus SNOWPACK. Since some observations begin farther below the surface, in this figure, observed ρ_0 is defined as the uppermost density measurement that is within 0.1 m from the surface. The SNOWPACK ρ_0 is calculated over the same vertical segment as the SUMup observation. The CFM-GSFC uses a prescribed surface density of 350 kg m^{-3} (green vertical line), which falls near many of the observed surface densities. (b) Histogram of observed surface density with the mean represented by the black line. Also plotted is the CFM-GSFC surface density of 350 kg m^{-3}.

Fig. A3: caption: I’m not able to follow the anomaly calculation. Wouldn’t subtracting each year’s mean from the record lead to discontinuities in the time series? And wouldn’t it be easier to simply detrend the time series? Because this part is methodological a bit more complex (see also my comment to Table 3), it might even be worth to move this part to a separate section in *2. Methods and data*.

We thank the reviewer for pointing out the confusion in this caption and we have updated the text in Section 3.2 where we first mention the seasonal signal. The reason we choose to subtract each year's annual mean is to account for the differences in magnitude from year to year. Thus, we end up with the seasonal signal (i.e., anomalies that are independent of the annual mean). The means from one year to the next are not so drastically different that they lead to massive discontinuities.

In Section 3.2, we have rewritten the beginning of the seasonal signal paragraph as follows: ***To examine how the models represent the seasonal cycle in spatially-integrated FAC during the RCl, we subtract the annual means from each year to isolate the seasonal signal. We then fit a sine wave to the data (Fig. A3) and quantify a seasonal signal from the amplitude following methods from Ligtenberg et al. (2012).***

New references

Kuipers Munneke, P., Ligtenberg, S. R. M., Noël, B. P. Y., Howat, I. M., Box, J. E., Mosley-Thompson, E., McConnell, J. R., Steffen, K., Harper, J. T., Das, S. B., and van den Broeke, M. R (2015).: Elevation change of the Greenland Ice Sheet due to surface mass balance and firn processes, 1960– 2014, The Cryosphere, 9, 2009–2025, https://doi.org/10.5194/tc-9-2009-2015

References used in response

Global Modeling and Assimilation Office (GMAO): MERRA-2 tavg1_2d_int_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Vertically Integrated Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/Q5GVUVUIVGO7, type: dataset, 2015