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.

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. 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)?
- 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.)?

Point-comments

Content-related (text)
Line 10: For which time are these statements valid? 1980, 2020 or averaged over the 40 years?
L84: I would call this section "**Methods and data**" (because you also present the SUMup observations)
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?
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.
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?
Section 2.2: Could you specify which scheme for vertical water percolation is applied 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)
L135: Why do you perform the vertical interpolation only for CFM output (and not for SNOWPACK – which also has a fine grid spacing)?
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)?
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).
L174: Why do you use 100 m as a lower limit (and not e.g. 150 m – the spin-up depth of SNOWPACK)?
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)
L269: "no change" might be a bit too restrictive. Maybe better "only negligible changes"
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…
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...
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?

Typos, phrasing and stylistic comments
Line 6: …Community Firn Model (CFM)**,** to quantify…
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."
L117: "scheme use to" → "scheme use**d** to"
L195: "form**ed**"
L197: "in the surface" → "close to the surface"?
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."
L366: change "MERRA-2 grid point" to "MERRA-2 grid cell" (also later in the text)
L374: "in simulating observations" → "in reproducing observations"
L375: I would rephrase this sentence.
L443: I would rephrase this sentence.
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."

Figures and Tables
Figure 1: Adding degree symbols and N/E to the latitude/longitude coordinates would help the reader.
Figure 2: I would state relative biases in percentages like specified in Eq. (4)
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"
Figure 6: caption → what caused the missing data?

Table 3: How did you distinguish between detectable and undetectable signals?

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."

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*.

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