Summary and general comments

In this study, present and near-future firn conditions of the Eclipse Icefield (Yukon, Alaska) are assessed based on contemporary firn cores and numerical firn modelling efforts. Firn stratigraphy, density and temperature from three cores (and ground-penetrating radar measurements) are analysed and contextualised/compared with previous measurements from nearby sites. The Community Firn Model (CFM) is then set up optimally for this site by choosing the best air temperature data for firn profile spin up. Furthermore, the optimal pair of degree day factor (DDF) and surface density are estimated for this site. An ensemble of simulations (2024 – 2033) is then performed with the CFM with different prescribed levels of air temperature changes [-0.1 to +1.0°]. Results indicate that with ongoing warming, firn at 15 m depth will likely become temperate with the associated potential of firn aquifer formation. This transformation is critical for the Eclipse Icefield site, because it served as a location for deep ice core retrieval.

This is a very interesting study that links observational data (retrieved firn cores) with firn modelling. The study is generally well written and supported by high-quality illustrations. In its current form, its rather a bit lengthy and I suggest to shorten certain sections slightly (e.g., 4.1, 4.2 and 4.4). Additionally, the discussion/interpretation of the firn cores and the modelling part are occasionally a bit detached and could be stronger interlinked. I suggest publication after the following major and minor comments are addressed:

Major comments

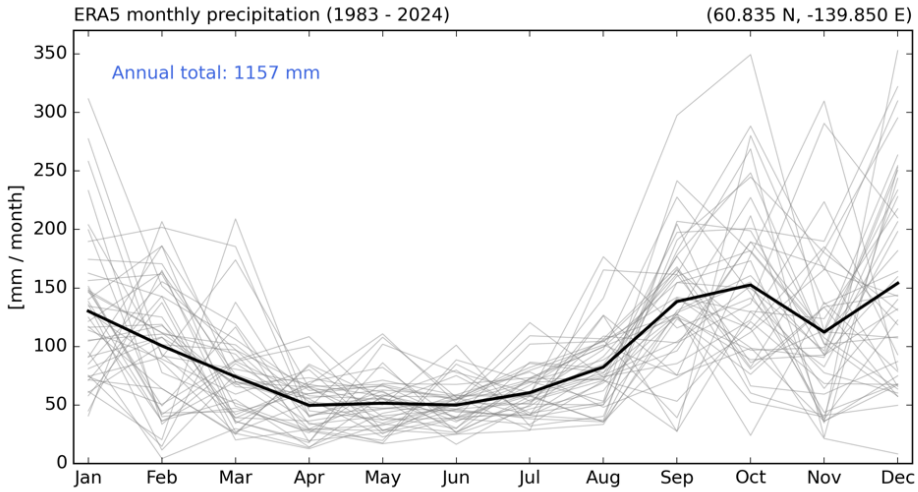
Improve section about CFM setup

I'm still a bit confused how you performed the sensitivity runs and the spin-up procedure exactly:

- In Appendix A, you nicely describe the sensitivity of the spin-up to the four different air temperature datasets. However, after only reading Sect. 2.3, I was uncertain if you exclusively use the firn profile generated with the NARR spin-up for all subsequent experiments or not. I would clearly state this in Sect. 2.3 and also briefly mention why you opted for the NARR spin-up.
- Just to be sure – my above assumption is correct, isn't it? You initialise all experiments starting in 2013 with the same firn profile?
- I'm confused which pair of DDF and surface density you use for spinning up the model (→ chicken-egg-problem ;-). As I understand from Fig. A2, you derived the optimal pair of DDF and surface density from the period 2013 – 2024. But which pair did you use to derived the initial firn profile for 2013 with the spin-up?

Assumption of time-invariant accumulation rate

Line 137: "accumulation rate of 1.4 m w.e. a^{-1} (McConnell, 2019) distributed evenly throughout the year". I wonder if this is a reasonable assumption. To check this, I briefly plotted the monthly precipitation of ERA5 for the closest grid cell:



The magnitude of precipitation in autumn/winter is a factor 2-3 larger than in summer, so there exists a distinctive seasonal cycle - of course under the assumption that precipitation from ERA5 (model resolution ~30 km) is representative for the Eclipse side. The annual total agrees well with the measured accumulation rate of 1.4 m w.e. a^{-1}. I think it would be worthwhile to check the sensitivity of the CFM results on seasonally variable precipitation.

Air temperature generator
I suggest to improve the description of the air temperature generator. Until reading appendix A, I was uncertain how you pool the daily air temperature data (you use daily means, right?) to compute the mean and standard deviation. It seems that you compute these statistics for every day of the year – correct? I would definitely mention this in the main text.
Furthermore, I wondered how realistic these synthetic air temperature time series actually are. Due to the random selection, there is probably a very high day-to-day fluctuation in air temperature. In reality however, air temperature might sometimes be more constant due to persistent weather patterns.
And finally, I'm astonished by the large spread of surface temperature (ca. 40° C) in Fig. 7b, which is probably related to the air temperature generator. Are the panels to the right showing the temperature for the last day of the simulation (2033-12-31)? Or is temperature averaged over a certain period?

Minor comments

Content-related
Line 29: I find the term "irreducible saturation" a bit odd. Maybe better "irreducible water"?
L30: How exactly do firn aquifer warm the firn?
L68: I don't understand this sentence: How was the plausible LLS depth (4.0 – 4.5 m) derived from the firn core observations?
L85: I'm unfamiliar with writing error/uncertainty propagation in this way. Do you have a reference for this equation? I'm also confused by the usage of "d" – I guess it is not used for an infinitesimal quantity because later on finite values are assigned to it (e.g., on L91: dL = dD = 0.25). Maybe it's better to replace "d" by the delta symbol?
L99: Why ± 0.2 m? Shouldn't it be ± 0.25 m (in accordance with lines 68/69?
L103: First I was confused about how the 12 and 1.5 hours fit together. But I assume you let the borehole equilibrate for 12 hours before you **start** installing/inserting the temperature sensors - right? Maybe you can write this more explicitly.
L108: Did you check that 15 s of equilibration time is sufficient (by checking that the measurements during the 30 s are approximately constant)?
L113: How is the uncertainty of 0.01° C selected? It seems to be somehow derived from the 15 s equilibrium time…
L126: Just out of curiosity: how was this semi-automatic picking performed?
L132: CFM provides multiple densification schemes, why did you choose the one from Kuipers Munneke et al. (2015)?
L132: I was uncertain what you mean by "assigned surface density" until I looked at table 2 and the following text. Maybe you could write here something like: "and a time-invariant surface density derived from a sensitivity test *(reference to later text)*"
L134: To which depth did you simulate firn in CFM? Which lower boundary condition for the heat equation was used? Dirichlet or Neuman?
L148: Although the sensitivity tests are explained in more detail in Appendix A, I would briefly mention some important facts here: over which time was the model run for the sensitivity tests? How was the firn profile initialised for the different sensitivity runs?
L153: How is spin-up time defined? The time required to refresh the entire simulated firn column?
L172: The division in section 2.4 and 2.5 is not entirely clear to me – maybe one could list all reference data in one section (e.g., as bullet points)
L262: How do you infer an ice thickness of only ~150 m from Fig. 6c?
L266: Explain what you mean by "reference model"
L277: Are the simulations shown in Fig. 9 just random examples from the 50 members? Or were they specifically selected?
L280: I don't understand this sentence, could you rephrase it?
L309: "and less than" → "and more than"?
L327: What do you mean by "peaks and cyclic variations"?
L389: I do not fully understand this sentence, could you rephrase it?
L418: How do you know that mean annual temperature at the two sites is virtually the same?

L421: I'm confused: B5 receives more solar radiation than B2 but shows nonetheless fewer melt features?
L444: "is < 2° C from supporting liquid water at depth" → could you rephrase that?
L453: Now I'm confused: What's the difference between number of PDD and number of melt events? It's probably helpful if you introduce and explain these metrics somewhere.
L469: "Additionally, the loss of firn pore space…" → difficult to understand, could you rephrase this?
L534: I'm not familiar with the Kruskal-Wallis test. Could you briefly explain what "H" represents?
Figure 6: Is the abbreviation "TWTT" explained somewhere?
Figure 9: Colorbar for temperature not very intuitive (transition from yellow to red/violet is normally interpreted as warming…)
Figure 11: Is the Colorbar of (a) identical to the one used for (b)? Because according to this colorbar, the north-eastern region seems very low elevated (~0 m) but it is higher in reality I guess…
Table 1: What limited the (different) bottom depths of the three recovered firn cores?
Figure A1: Why is there a range of simulation for the 4 experiments? Did you vary something else besides the four driving air temperature datasets?

Typos, phrasing and stylistic comments
L80: space missing between "kg" and "m^{-3}" (twice)
L146: "We tested a higher concentration of surface density values…" → "We refined the surface density spacing between 400 – 520 kg m^{-3} since…"
L146: space missing between "kg" and "m^{-3}"
L151: I would remove "to predict the evolution of the firn pack from 2024 – 2033"
L158: space missing between "C" and "a^{-1}"
L174: no space between "m" and "a.s.l." (same on line 176)
L186: "between the our" → "between our"
Figure A4: "…between 1979 and 2016 applied" → rephrase
Table B2: Decimal place errors in first column. E.g., "+0.05° C" → "+0.5° C"