Reviewer 1

I would like to thank the authors for the additional effort they put in the revision of the manuscript. Please find below some additional final remarks (line numbers refer to the revised manuscript with marked changes):

Day-to-day fluctuation in air temperature
Thanks for the additional graph in the replies comparing NARR, AWS and synthetic data. I'm a bit puzzled now about the distinctively different (and lower) fluctuation in the NARR data. Did you use monthly averaged air temperature data from NARR? If so, I would mention this somewhere…

Text edited to:

> *We spin the model up from ~1983–2013 (exact spinup time varies slightly among model runs as it is the time required to refresh the entire firn column and therefore dependent on densification rate and surface melt) using monthly averaged downscaled North American Regional Reanalysis (NARR) air temperatures for Eclipse from 1983 to 2013 Jarosch et al. 2012).*

L127: "we increase our uncertainty…" → still difficult to understand – I would rather remove this part. The following new sentence is also a bit hard to follow: Why is the borehole diameter (better use SI units to specify) related to the depth of measured temperatures?

The borehole diameter is not related to the depth of measured temperatures. Rather, both of those elements are related to our choice to ignore advective heat transport. Thanks for pointing out the confusion here; we've rearranged the text as follows to clarify:

> *We assign an uncertainty of 0.01°C to our temperature profiles at 20 cm increments when a 30 s equilibrium time was used. We ignore the effects of advective heat transport for three reasons: we were most concerned with temperatures below 10 m depth, boreholes were only ~5 cm across, and surface conditions were very similar on both days when temperature measurements were collected (sunny, light breeze).*

L157: Maybe better: "and a Neuman boundary condition for the heat equation at the bottom of the 50 m deep firn column."

Changed as suggested

L174: Very nice that you now consider seasonal variable snow accumulation. Could you show the monthly scalars you computed (respectively the seasonal variation in snow accumulation) in a plot (could be moved to the appendix)?

We have added the following figure:

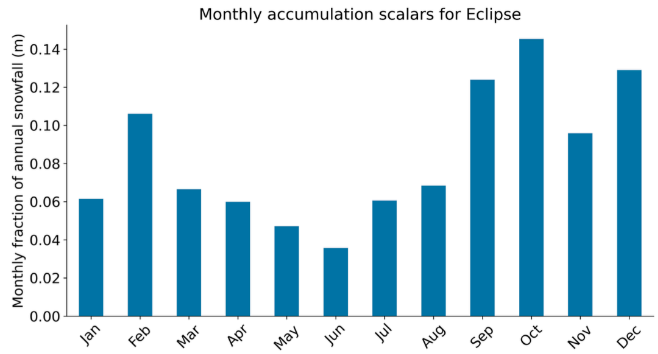

Figure B5. Monthly accumulation scalars calculated using an in situ snow accumulation record from Divide. We use four years of complete coverage (2004, 2005, 2006, 2008) to compute each month's mean fractional contribution to annual accumulation. We compute the fractional contribution to total positive surface change (ignoring ablation) to represent snowfall rather than net accumulation.

L490: Sorry, I'm still confused by your reasoning. I think I can (partially) follow why aspect and solar radiation could contribute to a slightly warmer temperature – but why does this configuration cause less melt in summer (B5 should then still receive more radiation – right)? If in doubt, I would rather remove this part because I think the meltwater percolation hypothesis is sufficient…

We included both hypotheses because each has a notable flaw. As you point out, it is unlikely for the higher solar radiation at B5 to be entirely a wintertime phenomenon, but Eclipse is high enough latitude that the sun position in the sky is considerably further south in the winter than summer, so this may play some role. The main flaw in the meltwater percolation hypothesis is the probably limited lateral transport of liquid meltwater in cold firn before it refreezes. We've modified the text as follows to more explicitly address the flaw in the first hypothesis that you point out. We do think both hypotheses are valuable to keep in the text because the meltwater percolation hypothesis doesn't seem sufficient to us.

> *One explanation for observed differences between sites B2 and B5 is that B5 on average receives slightly more solar radiation due to its southeast aspect, especially in the winter. Eclipse is high enough latitude that the sun position in the sky is considerably further south in the winter than summer (approximately a 30° difference in azimuth between noon on 21 June and noon on 21 December). This being a predominantly winter phenomenon could explain why B5 has fewer melt features despite being slightly warmer; however, we consider it unlikely that the wintertime southerly migration of sun position would outweigh the influence of summertime melt production at the two sites. Moreover, in areas with surface melt and percolation into the subsurface, the role of conduction in downward heat transport is comparatively minor relative to that of latent heat associated with the refreezing of meltwater (Cuffey and Paterson, 2010).*

L542: Forgot to include rephrased sentence: Additionally, firn loses pore space in response to warming more readily than it gains pore space in response to cooling; observed densification of the firn to date therefore has long-term consequences for runoff buffering (Thompson-Munson et al., 2024). Overall, I'm still a bit puzzled by this statement: I guess the rate of firn pore space increase during cooling is, after reaching sufficiently cold temperatures, mainly a function of snow accumulation, which is not mention here…

The rephrased sentence has now been included! Yes, we would agree that new snow accumulation would likely increase pore space if it can be retained without melting, "resetting" some of the firn column. However, the paper we reference here presents idealized warming and cooling experiments on Greenland's firn, focusing on the effects of air temperature on compaction and melt. We only mention it in passing to point out that the effects of firn densification in response to atmospheric warming probably aren't easily reversible and should be taken seriously.

L610: Spell out abbreviation "ANOVA"

Done

Fig. A4: Forgot to include rephrased sentence: Panel (b) shows model runs spun up with temperature values randomly selected from a Gaussian distribution based on elevation-corrected Divide AWS data. A historical warming rate of 0.024°C a−1 between 1979 and 2016 was applied to these data (Williamson et al., 2020).

The rephrased sentence has now been included

Typos, phrasing and stylistic comments
L14: "would represent likely indicate" à would likely indicate"

Done

L76: "we're" → "we are"

Done

L88: "We removed outliers…" → this sentence reads odd and should be rephrased

Rephrased to:

We removed outliers for all depths below the last summer surface if their density was > 917 kg m^{-3} or < 300 kg m^{-3} within uncertainty.

L154: Better (?): "We do not expect that using a different firn"

Changed as suggested

L201: “We” should remain in the sentence.

Double checked that “we” is in the sentence

L455: “can be observed Eclipse at the kilometre scale…” → fix’’

Changed to: “can be observed at Eclipse…”

Reviewer 2

I would like to thank the authors for engaging seriously with the reviews from both me and the other reviewer. I think the paper has been improved and would make a good addition to The Cryosphere. Many of my comments are stylistic or technical; however, I have three main comments that I think should be considered before the manuscript is ready for publication.

Main Comments:
[1] I would encourage the authors to define “extreme melt events” or “extreme individual melt events”. It appears the author’s definition of “extreme” is “high-intensity”. What is the intensity threshold that makes it extreme? It may be better to just say “intense melt events” rather than extreme, as in Greenland, extreme melt events commonly refer to the melt extent.

We have changed our phrasing throughout to avoid confusion with extreme melt events in Greenland. You are correct in that we are talking about high-intensity melt events. Specifically, we’re talking about an increase in the average individual melt event magnitude (mm melt produced). We have changed our language as follows to clarify this:

> *We suggest that more and higher-magnitude melt events during the height of summer combined with warmer wintertime temperatures promote the development of year-round temperate firn in the St. Elias. Model results for Eclipse show the development of year-round temperate firn at 15 m depth associated with an increase in total PDDs throughout the melt season, as well as with a greater number of individual melt events, higher average melt event magnitude (mm melt produced), and warmer winter temperatures, rather than an earlier or prolonged melt season (Fig. 8; Tables 4, 5). In Greenland, "extreme melt events" have been related to firn’s multi-year response to surface melt via the formation of thick ice slabs and ice layer complexes, which cause a near-surface barrier to downward percolation (Culberg et al., 2021). In the St. Elias, however, an increase in the number of melt events and in average individual melt event magnitude are more likely to result in sustained heat transport to depth because of the insulating effect of the region’s high annual accumulation (1.4 m w.e. a^{-1} at Eclipse) relative to accumulation rates in Greenland (0.3 – 1.2 m w.e. a^{-1}; Hawley et al., 2020; Montgomery et al., 2020; Burgess et al., 2010).*

[2] In the abstract the authors state, "…the development of year-round deep temperate firn at Eclipse Icefield is promoted by an increase in extreme individual melt events, rather than a greater number of small melt events or a prolonged melt season" (L6-8). I still am a bit unclear on how the authors arrive at that conclusion. Across most years, the median PDDs, number of melt events, and melt event magnitude are all significantly greater in model runs that produce temperate firn (Table B2), so why zero-in on the extreme individual events? How do the authors know that is driving firn warming more than the number of melt events?

Good catch! This is a holdover error that should have been corrected with the revision of the tables in Appendix B (now in the main text). The abstract text has been rephrased as follows:

> *In particular, the development of year-round deep temperate firn at Eclipse Icefield is promoted by an increase in the number of individual melt events and in average melt event magnitude combined with warmer wintertime temperatures, rather than an earlier or prolonged melt season.*

I think Figure 10 could be improved by making this a 3-panel plot (although I'm happy to consider other changes to the plot that the authors see fit). Similarly to how PDDs are displayed currently, additional panels could show the distribution of number of melt days between models that produce temperate firn and those that don't, as well as the distribution of melt intensity per melt event between models that produce temperate firn and those that don't. Since these are the three significant drivers producing temperate firn, it would be nice to see them displayed in a figure, and it may allow the authors to make their point clearer about why more intense melt is the main driver.

Figure 10 has been converted into the following 4-panel plot (we included winter temps in addition to you three suggested panels):

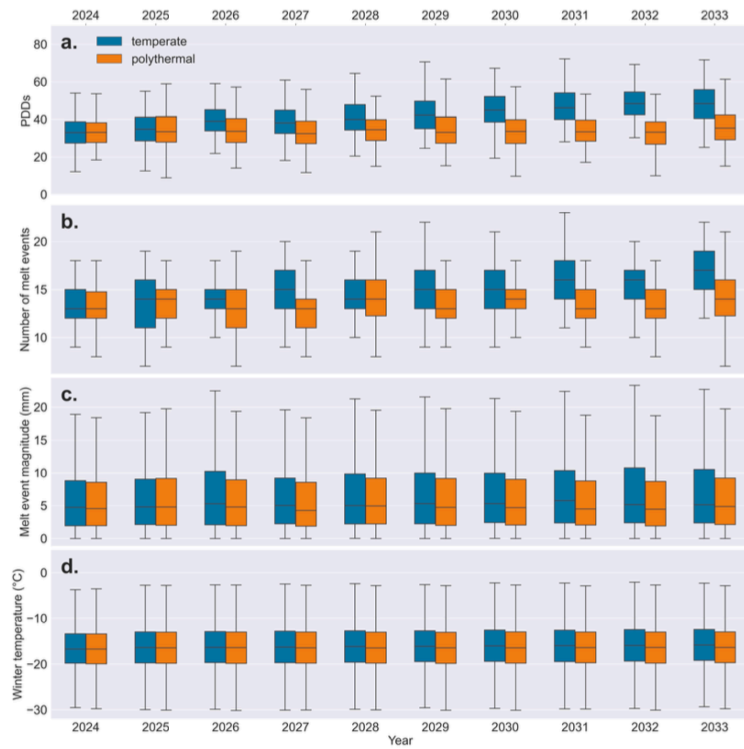

Figure 8. Melt season positive degree days (PDDs, a), number of melt events (b), melt event magnitude (c), and wintertime temperatures (d) each year for model runs that produce year-round temperate firn at 15 m depth by 2033 (blue) and those that do not (orange). We define the melt season here as May-September, and wintertime as October-March. Outliers are excluded for clarity.

As it stands, the authors state their conclusions slightly differently throughout the text. Just to highlight, on L457-459 in the Discussion, the authors state: "Model results for Eclipse show the development of year-round temperate firn at 15 m depth associated with an increase in total PDDs throughout the melt season, as well as with a greater number and more extreme melt events, rather than an earlier or prolonged melt season". That is slightly different from the abstract, though better supported by the statistics. Lastly, in the Conclusion on L489-490, the authors state "Development of year-round temperate firn at Eclipse is associated with an increase in total PDDs throughout the melt season and more extreme individual melt events rather than a greater number of melt events or prolonged melt season." This appears to be more similar to what is stated in the introduction. I would try to be as consistent as possible.

The conclusion text has been edited to be consistent with the updated abstract and discussion as follows:

> *Development of year-round temperate firn at Eclipse is associated with more total PDDs throughout the melt season, more individual melt events, and a higher average melt*

event magnitude combined with warmer wintertime temperatures, rather than with an earlier or prolonged melt season.

[3] I appreciate the revision of the tables in Appendix B. The editor and authors may feel this is unnecessary, but I am partial to reporting the true p-values to the readers. P-values that are significant below the author's confidence threshold could be highlighted or in bold text. Also, it may be nice to report the statistics being evaluated (e.g. difference in median PDDs, difference in median number of melt events, and difference in median melt event magnitude). It may make the authors conclusions clearer and less "hidden" behind the statistics. I would also consider moving these to the main text since they directly support the authors' conclusions. I think there are some figures/tables that could be moved to the appendix to make space (e.g., Figure 3, Table 1, Table 2).

The tables have been changed to those shown below, and moved to the main text.

Table 3. Difference in mean winter (October through March) temperatures (T, independent T-test) between model runs that produce year-round temperate firn by 2033 and those that do not. Significant results (at the 95% confidence level) are bolded.

Year	Temperate mean (°C)	Polythermal mean (°C)	T	p
2024	-16.63	-16.66	0.77	0.44
2025	-16.46	-16.53	1.49	0.13
2026	**-16.40**	**-15.54**	**2.84**	$\mathbf{4.5 \times 10^{-3}}$
2027	**-16.35**	**-16.51**	**3.34**	$\mathbf{8.3 \times 10^{-4}}$
2028	**-16.27**	**-16.55**	**5.89**	$\mathbf{4.0 \times 10^{-9}}$
2029	**-16.20**	**-16.56**	**7.48**	$\mathbf{7.6 \times 10^{-14}}$
2030	**-16.08**	**-16.50**	**8.87**	$\mathbf{7.6 \times 10^{-19}}$
2031	**-16.07**	**-16.49**	**8.78**	$\mathbf{1.6 \times 10^{-18}}$
2032	**-16.01**	**-16.50**	**10.23**	$\mathbf{1.5 \times 10^{-24}}$
2033	**-15.88**	**-16.44**	**11.94**	$\mathbf{8.0 \times 10^{-33}}$

Table 4. Difference in median number and magnitude of melt events (W, Wilcoxon Rank Sum tests) between model runs that produce year-round temperate firn by 2033 and those that do not. Significant results (at the 95% confidence level) are bolded.

	Median PDDs		Number of melt events		Melt event magnitude	
Year	W	p	W	p	W	p
2024	-0.13	0.90	0.02	0.98	0.27	0.78
2025	-0.10	0.92	0.81	0.442	4.2×10^{-3}	1.0
2026	**4.68**	$\mathbf{2.9 \times 10^{-6}}$	1.46	0.14	**2.74**	$\mathbf{6.1 \times 10^{-3}}$
2027	**4.73**	$\mathbf{2.2 \times 10^{-6}}$	**3.33**	$\mathbf{8.6 \times 10^{-4}}$	**3.03**	$\mathbf{2.4 \times 10^{-3}}$
2028	**4.72**	$\mathbf{2.3 \times 10^{-6}}$	0.93	0.35	1.17	0.24
2029	**6.01**	$\mathbf{1.7 \times 10^{-9}}$	**3.15**	$\mathbf{1.6 \times 10^{-3}}$	**3.06**	$\mathbf{2.2 \times 10^{-3}}$
2030	**7.63**	$\mathbf{2.3 \times 10^{-14}}$	**2.46**	**0.01**	**3.81**	$\mathbf{1.4 \times 10^{-4}}$
2031	**8.87**	$\mathbf{7.6 \times 10^{-19}}$	**4.87**	$\mathbf{1.1 \times 10^{-6}}$	**5.47**	$\mathbf{4.6 \times 10^{-8}}$
2032	**9.16**	$\mathbf{5.3 \times 10^{-20}}$	**4.21**	$\mathbf{2.6 \times 10^{-5}}$	**5.45**	$\mathbf{5.1 \times 10^{-8}}$
2033	**8.26**	$\mathbf{1.4 \times 10^{-16}}$	**4.37**	$\mathbf{1.2 \times 10^{-5}}$	**3.21**	$\mathbf{1.3 \times 10^{-3}}$

Table 5. Difference in median melt season timing and duration (W, Wilcoxon Rank Sum tests) between model runs that produce year-round temperate firn by 2033 and those that do not. No years showed significant differences between the medians at the 95% confidence level.

	Melt season start		Melt season end		Melt season length	
Year	W	p	W	p	W	p
2024	-0.18	0.86	0.87	0.38	0.87	0.39
2025	0.24	0.81	0.89	0.37	-0.15	0.88
2026	1.27	0.21	-1.65	0.99	-1.66	0.10
2027	1.36	0.17	-1.90	0.06	-1.72	0.09
2028	0.90	0.37	0.34	0.74	-0.44	0.66
2029	-0.02	0.98	-0.67	0.50	-1.32	0.19
2030	-0.21	0.83	1.55 p	0.12	1.50	0.13
2031	0.41	0.68	-1.15	0.25	-1.55	0.12
2032	1.20	0.23	1.06	0.29	-0.36	0.72
2033	-0.74	0.46	1.08	0.28	1.37	0.17

Minor Comments

L14: Remove the word "represent" where it says "… would represent likely indicate…"

Done

L30: "Firn aquifers account for much of observed firn water storage"… Much seems a little vague here. Maybe the authors could provide an estimate? An approximate water storage volume or areal extent of firn aquifers compared to a melt area extent? Not a big deal but could strengthen the sentence.

We've edited the text as follows to provide some quantitative grounding:

> *Firn aquifers can store large amounts of liquid water, and can retain water for several years, both delaying runoff and warming the firn (Ochwat et al., 2021; Miège et al., 2016; Jansson et al., 2003; Schneider, 1999; Fountain, 1989). For example, firn aquifers across Greenland have been estimated to store 140 ± 20 Gt of liquid water, buffering 0.4 mm of sea level rise (Koenig et al., 2014).*

Upon rereading the manuscript, I'm wondering if in the paragraph from L176-188 along with Table 2, it may be appropriate to move to the appendix since it is more detailed sensitivity tests that, while interesting and useful, could be available for interested readers at the end, which would shorten some of the main text. I think it also will keep more focus on running CFM under a suite of climate scenarios as the authors describe in the following paragraph.

Done

Section 3.1. I would recommend including a sentence in the beginning paragraph to highlight what the authors would like for readers to take away from this. It gets quite dense when describing the detailed stratigraphy in each paragraph. Maybe just highlight the key point for readers to have something to hold onto… maybe indicating that these measurements are important to demonstrate the substantial variability between cores, even spaced less than 1 m apart?

We have changed the first sentence to the following:

> *The stratigraphy of all three 2023 cores shows ice layers, ice lenses, and melt-affected firn throughout the core; however, variability among individual melt features is high among all three cores despite the proximity (<1 m) of cores B501 and B502 230 (Fig. 2).*

Section 3.2. A similar suggestion as above. The authors may even be able to start with the sentences: "In general, density increases with depth throughout the core. However, cyclic variations can be seen, which are likely seasonal, particularly in the top 10 m. Individual ice layers can also be identified by peaks in density" and reference Figure 4. I think it is helpful to know what the authors want the reader to see in the figure when referencing it. As it stands, the initial sentence to start 3.2 does not really provide any information for this section.

We've rearranged the text according to your suggestion, beginning the paragraph with:

> *In general, density increases with depth throughout the core. However, cyclic variations can be seen, which are likely seasonal, particularly in the top 10 m. Individual ice layers can also be identified by peaks in density (Fig. 3).*

Line 298: Where the authors say “at least some” could they just say how many runs?

We could, but it gets a bit unwieldy to report the number of runs for every depth just to make the point that the number is non-zero. We chose not to report for each depth here, so that when we do report specific numbers for each climate scenario later on, it helps emphasize that’s the bigger point.

Line 327: Show the same results as in 2016? Maybe clarify this in the sentence.

Line 327 states “Conditions associated with the development of year-round temperate firn at 15 m depth include lower mean winter temperature, higher total melt season positive degree days (PDDs), more individual melt events, and higher average melt event intensity.”

We’re unsure where the reference to 2016 is coming from. Perhaps the line number is a typo?

Technical corrections

Please ensure that the colour schemes used in your maps and charts allow readers with colour vision deficiencies to correctly interpret your findings. Please check your figures using the Coblis – Color Blindness Simulator (https://www.color-blindness.com/coblis-color-blindness-simulator/) and revise the colour schemes accordingly. --> Figs. 7, 8, A2, and A3

We have used the color blindness simulator to check all of our figures, including the ones mentioned here. Because there are fifty individual lines in these panels, it is impossible to easily distinguish all of them, even without any form of color blindness. However, the point here is to show the distribution and general range of variability among model runs, and these elements can be discerned.