Dear Dr. Etienne Berthier,
Thank you for the extension once again and allowing us the time to implement significant changes into the manuscript. We have included a track-changes document and a point by point description of the changes (below). We have added significant new material regarding the firn modelling and addressed comments about the spatial variability – among other things. We have also added supplemental information in order to go into greater detail on the modelling. We have uploaded a video supplement but it is still waiting to be given a DOI. Lastly, we have changed the title to "Evolution of the firn pack of Kaskawulsh Glacier, Yukon: meltwater effects, densification, and the development of a perennial firn aquifer" to better match the new manuscript. We believe these changes will make the manuscript much stronger.

Thank you,
Naomi Ochwat, on behalf of the coauthors.

Below you will find the description of the edits the reviewers requested and how we responded to them.

Reviewer 1

General Remarks

The study by Ochwat et al. has been modified substantially. I would like to thank the authors for making a thorough effort in revising their study. Generally, I believe the study has improved and has gained in focus. The inclusion of firn modelling supports the main message that the aquifer is likely a new phenomenon (but also raises some questions, see below under detailed remarks).

My main concern is that the authors compare firn cores that are as much as ~10 km apart (the 1964 Divide core and the 2018 Kaskawulsh core, the distance between the 1964 and 2006 Divide cores is much smaller but unknown), while not commenting on potential issues of such a comparison. I am concerned with this point based on my own research experience: I was involved in a study quantifying firn changes from 1973 to 2018 on Abramov glacier, Kyrgyzstan, Central Asia (the paper will be published in Journal of Glaciology within the next couple of weeks). When we visited the glacier in winter 2018, we thought that we knew the exact location of the historic 1973 firn profiles and drilled at that site. When back from the field, we received further historic documents which showed us that we had drilled about 250 m away from the historic site. We visited the glacier again in summer 2018 and drilled at the historic location. Together with extended GPR measurements, this provided us with the possibility to quantify short-scale variability of accumulation rates and firn properties on the relatively large mountain glacier (24 km2). Although both drill sites look very similar (they both seem so be located in the same flat plain), the variability is large. Over a distance of 250 m, mean annual accumulation varies by a factor of almost 1.7. If we had compared the historic 1973 profile to our core drilled 250 m away from the historic site, our conclusions with respect to firn changes (accumulation, firn ice content, and more) would have been dramatically wrong. I am aware that Abramov glacier probably shows more small-scale variability as the accumulation area of the much larger Kaskawulsh. But even on the Greenland ice sheet, I would be cautious when comparing cores that are ~10 km apart. Hence, I would like to ask the authors to at least mention such potential issues and to thoroughly evaluate and discuss to what degree drill sites, conditions at the drill sites and the cores are comparable.

This is an interesting point, and we respect these concerns. We feel comfortable with the general climatic homogeneity of the plateau/divide region of the Kaskawulsh Glacier, given what we know about it, but agree that this requires appropriate qualification and caveats. Our own two cores, about 1 m apart, have considerable stratigraphic variability, despite similar winter snow depths and average firn densities, so we appreciate the complexity of spatial variability with both accumulation rates and melt-affected firn.

Our comfort comes partly from the relatively flat and uniform character of the plateau. There are nunataks, but the elevation of the glacier ranges by less than 200 m (2550 ± 100 m) over an area of more than 63 km^2, and available snow accumulation data from our site, the Copland site, and the IRRP A 1960s core sites are all consistent, with average values in the range of 1.6 to 1.8 m w.e. yr^{-1}. The latter number, 1.8 m w.e., is close to what we measured for the 2017-2018 winter accumulation at our core site and based on isotopic data from within the core (three consecutive winter (negative) peaks that were not washed out). This is also the average value found from 7 years of measurements at the Copland site, about 12 km away. So we agree that there can be strong horizontal variations in snow accumulation, and have seen this, but the available data support that accumulation rates are similar where it has been measured on the upper Kaskawulsh. The value we use, 1.8 m w.e., is partially informed by data at the core site, as well as the regional data. We do include some uncertainty analysis associated with this value, and have now added some discussion concerning potential issues of spatial variation, L. 89-94

Detailed Remarks

Lines 315-319: I do not agree to these statements. The fact that the firn was at 0 °C in 2006 (0 to 10 m depth) and cold between ~1 m and ~15 m in 1964 does not proof firn warming. To reliably quantify firn temperature changes, temperatures below the depth of zero annual amplitude (roughly 10 to 15 m depth) need to be compared. There is a high risk that the differences discussed here and shown in Fig. 5 are just the result of weather conditions and not a climatological signal. This risk is amplified (i) by comparing average temperatures for only the top 10 m and (ii) by the 1964 core showing 0 °C at ~15 m depth, potentially indicating that the firn was temperate already then. At the minimum it needs to be acknowledged that the data have to be used with care when quantifying firn temperature changes. I also suggest evaluating and discussing potential uncertainties.

Thanks for raising this, we totally agree. Temperatures and densities in the seasonal snowpack and near-surface firn cannot be compared from one year to another, as they are recording weather rather than long-term trends. Whether it was a cold winter or warm spring, etc. We have now removed this from the results and discussion, including Figure 5. Depths greater than 10 m are safer, as these are below the annual air temperature wave (via thermal diffusion), but the ‘textbook’ understanding that 10-m temperatures reflect mean annual conditions also does not apply where there is extensive meltwater refreezing (latent heat release), and where this can vary a lot from year to year. This is well illustrated in the Figures in the new Supplemental Information, and also by the observation of deep temperate firn at this site, despite mean annual air temperatures of about –11°C.

The second point, that Grew and Mellor (1966) reported temperate firn at 15-m depth in 1964, is particularly important, and we apologize that we missed this. We divided the task of revising the manuscript and the person charged with the modelling effort (SM) was aware of the Grew and Mellor core from the Icefield Ranges Research Program reports, but not of this particular CRREL technical paper, or the borehole temperature measurements. We therefore did not present a consistent story of the firn warming, and had not realized that the initial conditions in the firn model were inconsistent with this observation. An embarrassing error on our part. We now acknowledge this discrepancy, and perform sensitivity tests on the initial conditions in the model to try to better understand the model disagreement. We still present a similar ‘reference model’ for the 55-year firn temperature evolution, which predicts significant firn warming and the recent development of the firn aquifer, but are much more cautious in our interpretation. This reference model uses the bias-adjusted ERA climatology and the firn model parameters as calibrated at DYE-2 in Greenland (Samimi et al., 2020), but is not locally calibrated or validated, as the reviewer notes, so we cannot be too confident in either this model or the climatological forcing.

That said, we model strong fluctuations in the 10- and 20-m temperatures through the period 1965-2019, and even in the case where the model is forced to produce temperate deep firn in the spin-up simulation (Figures 7, 8, and S4), the firn refreezes in the 1970s and experiences decadal-scale warming trends, not returning to temperate conditions until ~2015. We discuss these deep-firn temperature trends from the model, but are now more careful in arguing for firn warming at this site, making clear that this is speculative and based on a model result. The inference that firn has warmed at this site is now removed from the title, abstract, and conclusions.

Lines 351-352: Meltwater which does not ...” I do not understand the statement, does the model also simulate lateral drainage? If yes, consider updating the model description. If not, please remove the statements as this would not be a result of your modelling efforts.

This is correct, thank you – we rewrote this sentence, Lines 381-382. In the model, water that percolates to the base of the firn column (35 m in the model re-runs for the revisions) is assumed to leave the system, so we refer to this as ‘runoff’ or ‘drainage’, but we don’t explicitly model lateral runoff. See Line 381-383 and Lines 589-591.

Lines 375: These Characteristics ...” a bit confusing, the previous sentence describes the current situation, not the original situation.

This sentence has been deleted because we appreciate we cannot say for certain that all of the characteristics listed in the previous sentence have changed since 1964.

Lines 379-380: As mentioned above, I do not think the data fully support this conclusion.

The density results support this conclusion, but we agree, not the temperature data. We have removed the last part of the sentence "and temperature". This section has been extensively rewritten, Lines 449-462. We still discuss the possibility of firn warming, but note that this is just a model result and we don't have confidence in it.

Lines 388-394: This could be placed in the introduction or description of data and methods. In my opinion, these are general statements based on the literature and do not fully fit here in the discussion.

This is a good comment; it's true, this was formulated as more of a literature review. We reworded and shortened this discussion throughout this paragraph, now Lines 464-474. We retain some discussion of the processes as part of the interpretation of our density structure.

Line 397: "... affects continue ...", something is wrong here.

This has been addressed in the rewritten text, Line 469.

Lines 399: I do not understand these comments. Sorge's Law has been derived from the study of dry firn at Eismitte, roughly at 3050 m near the centre of the Greenland ice sheet (*Sorge, 1935*). It is intended to reflect certain basic characteristics of dry firn under a constant climate. It was not intended to be valid for firn which experiences substantial melt under a warming climate. To my understanding, this is also how the law is formulated and the term "Sorge's Law" coined in *Bader* (1954).

The discussion concerning these two points has been reformulated: "However, with increasing meltwater percolation and refreezing effects, higher densities are common in the upper portions of the firn, as observed in our cores," Lines 469-470. We use the discussion to explain that the firn on Kaskawulsh does not follow Sorge's Law for dry densification, though perhaps this is too obvious. We shortened and simplified this section, but still make some comments on the complex density structure caused by the meltwater refreezing.

Lines 437-438: My apologies for not making the Central Asia glacier studies available that document a firn aquifer on a mountain glacier already in the 1970s. Please find the studies by *Glazirin et al.* (1977) and *Kislov* (1982) available for download under this link: https://drive.switch.ch/index.php/s/51wrYzVb9r4XRSh

This sentence "Apart from these studies, there have been no other published reports of PFAs on mountain glaciers (Christianson et al, 2015)" has been removed. We have also added a few sentences about other PFAs found, one on Abramov glacier in Russia by Glazyrin et al., 1977 (translated by Eduard Khachatrian) and another on Austfonna Ice Cap

(Zagorodnov et al., 2006), Lines 531-535. Unfortunately, we were not able to translate the PhD thesis.

Lines 489-490: This is a conservative estimate ...” I do not fully understand what is meant here.

This has been reworded to “this estimate of thinning is likely low” on Line 603.

Figure 1: I could not find a clear reference to the IRRP A Site in the text. I suggest adding a clear reference to this site in the text (what, when and who measured there) or remove the site from the figure.

Thanks for pointing out this omission. This is the IRRP site for the earlier snow-pit work and the 15-m borehole. In the IRRP there are several sites that range in alphabetical name (A, B, etc). An additional sentence has been added to the caption “IRRP A site is the site of the 1964 firn core (Grew and Mellor, 1966).” We also rewrote the study area section to better explain this, Lines 87-91.

Figure 6: Which location is modelled? While a location is clearly indicated in, e.g., Figure 5, there is no information here and also in the text where Figure 6 gets referenced. On line 233 it says that “the study site” is modelled, furthermore it is stated that model forcing ERA data are compared to the “Divide” meteorological observations. However, the study refers to a rather large area with locations at different elevations. Hence the question what exactly is modelled?

This is intended to be a model of the core site, at an elevation of 2640 m in the upper Kaskawulsh accumulation area, but in truth there is nothing in the model that is specific to that point (60.78°N, 139.63°W), other than the elevation for the bias-adjustments in the meteorological forcing. Rather, the model should be interpreted as representative of general conditions (the glaciological and climatological setting) of the upper plateau and Kaskawulsh-Hubbard divide region. The ERA5 reanalysis uses 0.25° grid calls, ca. 28 km, so our climate forcing is a general regional representation, not specific to a particular point. We discuss this now, Lines 262-269.

Figure 6: Related to my comment above: It is not fully clear what is modelled and whether the model output can be compared to the Divide, the Kaskawulsh field data or both? Nevertheless, I note that the modelled firn temperatures at 10 and 20 m depth in the 1960s are around -12 °C while the 1964 core drilled at Divide indicates a temperate firn regime (0 °C at ~15 m depth). I consider this a substantial disagreement between measurements and model.

Agreed that this is a substantial disagreement, which we now discuss directly, and as discussed above. But in terms of the modelling, these sites (IRRP A, the Copland weather station, our core site, the Copland camp) are all within the same ERA5 grid cell, so our meteorological forcing cannot be compared with AWS data or something that is specific to a point. It is at best a representation of the regional climatology. It is our assumption that similar elevations within the accumulation area of the broad plateau region will experience similar meteorological conditions. We explicitly state this assumption now, Lines 262-269.

Figure 7: It looks like there are white areas at the top of Figure 7a (data gaps or variations in surface elevation?) which do not appear in Figure 7b. Why the difference?

Apologies, these were 'off-scale' (below -20°C) temperatures that saturated in the contour plot and were rendered white. Corrected in the revised Figure 6.

References: Now the Machguth et al. citations are fully confused. There is a 2016 paper in Nature Climate Change and a 2006 paper in Geophysical Research Letters. You have now created a combined citation of both papers :-) Please correct this citation but also check all the other references for correctness.

Apologies for the confusion. We have corrected the references and double-checked everything again.

References not listed in the manuscript

Glazyrin G.E., Glazyrina E.L., Kislov B.V. and Pertzinger F.I. (1977) Water level regime in deep firn pits on Abramov glacier [in Russian], volume 45. Gidrometeoizdat

Kislov, B.V. (1982) Formation and regime of the firn-ice stratum of a mountain glacier [in Russian]. *Ph.D.thesis,* SANIGMI Tashkent.

Sorge, E. (1935). Glaziologische Unterzuchungen in Eismitte. Wissenschaftliche Ergebnisse der Deutchen Gronland-Expedition Alfred-Wegener 1929 und 1930-1931, 3, 270. in: K. Wegener, im Auftrag der Notgemeinschaft der Deutschen Wissenschaft (Ed.), Band III, Glaziologie, 1935.

We have added Glazyrin et al. (1977) and Sorge (1935), but we could not find a translator for the PhD thesis.

Reviewer 3

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

The manuscript describes firn evolution at a high elevation site on Kaskawulsh Glacier, St.

Elias Mountains, Yukon. This is a highly relevant topic and the presented results contribute to an improved understanding of ongoing trends in firn density, temperature and potential development of firn aquifers in mountainous environments. Generally, I find that the observational datasets are well described and interpreted. Also, the comparison with model output is valuable. Still, I have some moderate to major concerns that I would like to see addressed, primarily related to the lack of model calibration and the interpretation of surface lowering. If it would be an option to perform some additional model experiments, I would highly recommend to perform some additional runs to calibrate melt rates e.g. by minimizing the misfit between modelled and observed subsurface temperatures. Right now, discrepancies between the model and observations are hardly discussed and the lack of model calibration makes it difficult to draw strong conclusions on trends in firn conditions. My specific comments are listed below.

Specific comments

L39-41: This needs to be reformulated. The phrase "If the surface continues to melt" should be removed, since refreezing will happen directly when melt water enters cold snow/firn. 'Warming firn' does not necessarily lead to more refreezing, rather the opposite.

Reworded as suggested, Line 40.

L42: A firn aquifer only forms if the water does not directly run off through moulins/crevasses.

Agreed; we added a short comment to note this, Lines 42-43.

L42-44: It would be good to consistently use "perennial firn aquifer" rather than "firn aquifer", since here long-term (multi-annual) storage of water in firn is meant.

Thank you, revised to reflect this suggestion and now "PFA" is defined and used throughout.

L51-52: Another useful reference for the Canadian Arctic is Noël et al. (2018; https://doi.org/10.1029/2017JF004304), and for Svalbard references to Van Pelt et al. (2019; https://doi.org/10.5194/tc-13-2259-2019) and Noël et al. (2020; https://doi.org/10.1038/s41467-020-18356-1) could be relevant to add.
L55: Here Machguth et al. (2016; https://doi.org/10.1038/nclimate2899) could be cited.

We have added these references as suggested.

L69: I suppose this refers to air temperature? Please clarify.

Yes, clarified as suggested Line 81.

L107-117: This part fits better at the start of the next section (3.2).

Rearranged as suggested, Lines 132-144.

L137: I suppose L_unc should be dL (or dL in the equation should be L_unc).

Thank you, yes, revised to *dL*, Line 156.

L186-190: "Surface lowering associated with refreezing" is confusing. Surface melting leads to thinning and gravitational settling of the snow as well, but refreezing just adds mass to the existing vertical column and does not (directly) cause any thinning. See also my later comment.

Yes this is true, that is what we intended to convey. Revised to " There is surface lowering associated with melting but without associated mass loss, due to subsurface refreezing", Line 205. Also revised to clarify this in Lines 211-212.

Section 3.4: It appears that no calibration of the model has been done, presumably because there were no melt observations to compare to (?). This currently makes it very hard to trust the model output, especially since there appear to be major biases in modelled subsurface temperatures, which may indicate an underestimation of melt rates. See also my later comment.

This is a fair comment, and we agree – lacking local energy balance, melt, or firn observations to constrain the modelling, it should not be interpreted as a rendering of reality. Rather, the modelling has been added to explore potential scenarios for the long-term firn evolution at the core site, and to examine questions of whether the PFA may be a recent or long-term feature. We agree that since the model is not observationally constrained, much more sensitivity analysis is needed here, and we should not be presenting just one model scenario as 'truth'. We have now added a range of model sensitivity experiments to a) perturbations in the climate forcing, b) sensitivity to model parameters, and c) sensitivity to the spin-up (initial) conditions. Most of this has been added to a new Supplementary Information section, but we also added new Figures 7 and 8 and a discussion of model sensitivity in the main text. The 'reference model', using the ERA climatology, and with model parameters from a calibration at DYE-2 in southern Greenland, is still presented as our 'default' scenario, but the model sensitivity experiments show that this reference-model firn evolution is not well-constrained. Significant biases in the climatology or different assumptions about irreducible water content, in particular, can lead to either permanently temperate or permanently polythermal firn at this site (permanent referring to the simulation period, 1965-2019). This is presented and discussed now, and we feel that the only robust conclusion is that the climatological and glaciological setting of the upper Kaskawulsh render it very close to the tipping point between polythermal and temperate firn.

It is hard to know whether there are major biases in the subsurface firn temperatures, as there are limited data. But the reviewer is correct, the only historical data we do have, from borehole air temperatures in a 15-m firn core collected in July, 1964, indicate temperate conditions at 15-m depth at that time (Grew and Mellor, 1966). Spot measurements of borehole air temperatures during the summer melt season are not strongly reliable, when meltwater can enter the borehole and mixing occurs with the surface air (e.g., wind-pumping, convection). The methods used for the 1964 borehole temperature measurements are not well documented in Grew and Mellow (1966), but temperature appears to have been logged at 1-m intervals over a short time (< 1 day), without evidence of capping the borehole or installing a thermistor string for an extended period (e.g., as discussed in Zagorodnov et al., 2006). However, taking these data at face value, there is certainly a chance that firn has been temperate since the 1960s (or longer) at this site. We discuss this now, and model experiments explore the changes in model settings that are necessary to produce temperate firn at this time (Figures 7 and 8). It requires: (i) a warming of at least 1.9°C in the ERA climate forcing, (ii) an increase of more than 36 W m^{-2} in the ERA incoming longwave radiation, (ii) an increase of at least 62 W m^{-2} in the ERA incoming shortwave radiation, or (iv) snow albedo values of less than 0.65 (vs. our estimate of 0.78, as a JJA mean value), or (v) some combination of the above. On their own, these represent strong anomalies or significant deviations from the reference model, particularly as the ERA temperature and radiation data have been bias-adjusted against regional observations, but a combination of these biases may certainly be possible. Low values of capillary water retention, e.g. irreducible water contents of less than 1.2% by volume, also facilitate deeper meltwater infiltration and firn warming to greater depths, but are not enough on their own to bring the firn to a temperate state. In short, we cannot rule out that the model underestimates melt extent and/or meltwater infiltration, and therefore has a cold bias in the firn temperatures. We acknowledge this now and present some of the model experiments discussed above. We discuss this further below in response to the reviewer's questions about model spin-up assumptions.

Section 3.4: Most likely the subsurface model also simulates density evolution, but this is not mentioned here and no graphs of it are shown in the results. It would be an important validation of the model results if it could be shown that simulated density evolution matches the observed densities well.

True, the firn model does simulate densification and the stratigraphic evolution of ice layers, though this has not been calibrated or constrained in the DYE-2, Greenland work as well as the coupled thermal and hydrological parameters have been evaluated, via thermistor measurements and TDR probes in firn pits (Samimi et al., 2020). We have added the modelled densities and densification rates to Table 2 and Figure 5. The increase since the 1960s is consistent with the observations, with average firn density (from 4- to 35-m depth) increasing from ~640 to ~680 kg m^{-3} from 1965 to 2019. The modelled firn density in 2018, the time of acquisition of the firn core, was 682 kg m^{-3}, compared with the observed value of 698 kg m^{-3}. This is now discussed, Lines 403-404.

Section 3.4: I am missing a description of how the model and in particular the subsurface conditions were initialized, i.e. if some spin up has been done.

Apologies, this should have been explained. This has been added to the manuscript and Supplemental Information in some detail now, with Figures 7 and 8 also illustrating the sensitivity to model spin-up assumptions. We start with linear temperature and density profiles and then do a 30-year spin-up with perpetual 1965 climatology, i.e., running the energy balance and firn model through the 1965 annual evolution 30 times. This develops the 35-m temperature, density, and ice-layer stratigraphy that is used for the initial conditions. There is no memory of the initial linear profiles, but the spun-up firn conditions certainly influence the 1965-2019 evolution, so this is another important source of uncertainty in the model simulations (Figures 7B and 7D; Supplemental Figure S3). That said, even when forced to temperate initial conditions in model experiments, the firn refreezes in the model in the 1970s and remains polythermal until the last decade (Figure 8). This may be due to a systematic underestimate in the melt rates or meltwater infiltration in the model, as discussed above, but it provides some support for the model-based inference of decadal-scale firn warming and the recent development of the PFA.

L249: "heat advection from melt water percolation" seems odd, since melt water typically is at 0 degrees C and if it encounters cold snow it will just refreeze thereby releasing heat. This is the only way in which heat is "advected", but that is probably not meant.

This is what we meant, but we did not explain this properly. We allow for rain that can be above 0°C to add to the meltwater, in principle, and consider this a source of heat advection if water a bit above 0°C percolates into sub-zero firn. First the water cools to 0°C, warming the surrounding firn (the heat advection), then it will refreeze and release latent heat. The latter is by far dominant and we have revised the text to focus just on refreezing here, Line 277.

L251-252: The symbols k_h and k_w are mixed.

Apologies, corrected, Lines 280-281.

L284-293: It remains unclear how melt-affected and not melt-affected firn are distinguished. I think this is an important aspect, because if there is indeed non melt-affected firn with a firn aquifer below, that probably implies that fast deep melt water percolation through piping is an important process here. I would like to see additional discussion on this in the manuscript.

The following sentence has been added Lines 112-114 "Melt-affected firn is distinguished by ice layers, ice lenses, or can be indicated by the lack of grain boundaries, the presence of air bubbles, and opacity." See also Lines 324-326 for added discussion on this in the results. Additionally, a photo has been added to the supplementary information, Figure S5.

L306-313: What is the density difference between 1964 and 2018 when considering the mean density between the LSS and 15m depth?

We have added a sentence on Lines 344-245: " The difference between the average densities from the upper 7 m in the 1964 and 2018 core is 33 kg m^{-3}, which is an increase of ~7%. " and we have elaborated on this section including information from the modelling results.

L318-319: This is comparing snapshots of subsurface temperature to extract trends. Since subsurface temperatures may vary strongly from year to year, care should be taken to determine long-term trends based on only two snapshots in time.

You are completely correct, we have removed this section of the manuscript, eliminating the temperature data from both the Divide site and the IRRP. The data cannot accurately be compared, as you note, as it is related to the recent weather and not long-term changes.

L344-345: These kind of statements about melt trends are hard to defend without any calibration or validation of melt estimates against observations.

This is fair, we have qualified this now to note that this is just based on the ERA climatology, and we lack in situ observational constraints. See Lines 361-362.

L375-376: "due to increased presence of ice layers": Is there any information on ice content in the 1964 core?

There is not any information on the ice content in the 1964 core, but Grew and Mellor (1964) discuss meltwater percolation and refreezing, and present stratigraphic plots that display some ice lenses, but with relatively low densities and no mention of ice layers. There is a figure in the report that displays ice lenses but only in the first several meters of the snow accumulation and not the firn (Grew and Mellor 1966, Figure 2).

L381-383: The modelled subsurface temperature trends may be reasonable, but the absolute values are quite a bit off when comparing temperatures in Fig. 5 and Fig. 7. This discrepancy is important to discuss in much more detail in the manuscript, especially since many of the conclusions for example on when firn became temperate and when the PFA may have formed are based on the modelled temperature evolution.

Agreed, now revised as discussed above and examined in some depth in the supplementary material. See also new Figures 7 and 8.

L382-383: "The ERA5 climate analysis": This is rather an analysis of snow model output. Please reformulate.

Thank you yes, this is from the energy balance and firn model as forced by ERA. Revised in the revised discussion of results, Lines 455-462.

L388: "increases" and "effect"

This section has been revised and typos have been corrected, Lines 464-474.

L391-394: I would suggest to reformulate this. It is unclear what this "first stage of densification" is. In my view there are two processes that affect densification 1) gravitational settling (which will go faster at higher temperatures) and 2) refreezing. Refreezing will increase subsurface temperatures, which in turn may increase the densification rate by gravitational settling/packing. That is a completely different sequence of processes than described in Line 391-394.

We have revised this section, Lines 464-474. There is also the process of the snow grains being rounded due to warming temperatures, which impacts the settling. We rewrote the line on 'first stage' to " Melting rounds snow grains and increases the rate of the first stage of densification ", Line 465.

L397: "effects"

Thank you, this section has been revised and typos have been corrected.

L411-412: "The firn model predicted the development of wet, temperate conditions in the deep firn following the 2013 melt season, although it took two years to fully develop (Figure 7)." But the observations reveal that the firn was already temperate in 2006. This should be acknowledged.

This was not really the cases – it was just a deep snowpit in 2006, extending to 7 m, and we have no knowledge about the firn temperatures below this. We have included more information on the modelling but have removed the temperature data from 2006 and the discussion of this data.

L440: "Kuipers Munneke et al. (2014)"

Thank you for catching this. We have gone through the manuscript and edited the reference throughout.

L445-446: Kuipers Munneke et al. (2014) indicate what accumulation and melt conditions favour the development of firn aquifers. So in addition to the accumulation comparison it would be good to also compare melt rates with rates observed in southeast Greenland.

Good suggestion. We have added this discussion, Lines 548-551. We include melt estimates from southeast Greenland according to firn aquifer studies in the area by Miege et al (2016) and Miller et al. (2020). Indeed, both the accumulation and melt regimes are very similar to those in southeast Greenland.

L453: Temperature of the firn will not have a major impact on the perennial firn aquifer. Typically once a perennial firn aquifer has formed the firn above it is temperate and the winter cold wave does not penetrate deep enough to case any refreezing. A factor that is important though is how easily the water can runoff via moulins and crevasses.

The first part is true in a warming climate, which we are currently experiencing, so we agree. However, sustained cool conditions can refreeze the deep firn from above, although on a diffusive (decadal) time scale vs. the potentially rapid work of meltwater infiltration and latent heat release. e.g., if the ~10-m temperatures cool off due to reduced meltwater and cooler air temperatures, these eventually cool and refreeze the underlying firn. There are examples in Figure 8 and in Figures S3 and S4. We revised the discussion in the main text and now discuss the deep firn temperatures and the possibilities of cooling in the Supplementary Material. Drainage in crevasses or moulins is discussed on Line 42-43 – very true that this is the best possible way to drain the firn aquifer in a flat-lying area like this. We don't have radar data or other evidence of such features, and the presence of the water table that we drilled into suggests a lack of such features.

L454-455: Internal accumulation commonly refers to the amount of refreezing below the last summer surface, which is probably not what is meant here. See for example Cogley et al. (2011; https://wgms.ch/downloads/Cogley_etal_2011.pdf).

Agreed, thank you for catching this. Actually, reading the definition in the link you provided verifies that this is what we meant. However, through this it was brought to our attention that we included the wrong Schneider reference; we mean Schneider and Jansson 2004, not Schneider 1999. To clarify, we separated the sentences, Lines 560-561.

L469-470: Ice layers in snow and firn happen in any accumulation zone that experiences some melt, which is the case for the vast majority of glaciers on Earth. Hence, the presence of ice layers in firn is not something special. Please reformulate.

Thank you, agreed and rephrased, Lines 575-578.

L474-475: This is an important notion. I understand that there may not be any melt observations to make use of, but I would instead strongly suggest to perform new modelling experiments where one or more parameters affecting the modelled melt rates are calibrated such that a best match between modelled and observed subsurface temperature is achieved. Right now, it seems that modelled melt rates are underestimated, which would result in too little water percolation and refreezing in snow and firn, thereby explaining the currently underestimated subsurface temperatures. With a calibrated model, confidence in modelled melt rates and firn conditions would considerably increase!

We have added sensitivity experiments that increase melt rates to produce temperate firn, although we are hesitant to trust this model scenario more than the 'reference model'. We have added these results now though, Figures 7 and 8, discussed on Lines 416-445. Of course it is true – they imply much more melt, drainage/ablation, and denser firn.

L477: "with most of the meltwater refreezing": It is unclear if this is still the case. The subsurface temperatures reveal that firn was already temperate in 2006 implying that already then some melt water did not refreeze.

We revised this discussion and added some additional numbers, Lines 587-591:
"Within the model, 96% of total meltwater refreezes over the 55-year simulation, but this is reduced to 86% for the period represented by the firn core, 2005-2017. The remaining 27% drains to the deep firn through this period, where it is either retained within the PFA or it may drain from the system. A total of ~1.3 m w.e. 'runs off' through the period 2005-2017. In the model, this drains through the bottom layer and leaves the system; in reality, this water may drain through lateral transport in the PFA or at the ice-firn interface."

Over the full period, some meltwater infiltrated to depth in warm summers with high melt, but it does not escape the system (drain to 35 m depth) until after 2017, when the deep firn becomes temperate. As noted above, we don't know that this was the case in 2006 – only that the seasonal snow and upper ~2 m of firn were temperate, with no knowledge of deep firn conditions. It is possible though, per Figures 7 and 8 and the discussion in the results.

L484: It would be nice to have an additional figure showing the modelled density evolution.

We have added this figure, please see Figure 5E.

L487: "0.73+/-0.23 m". If this is calculated from Eq 6 then, if I am correct, this is not the actual surface lowering, but rather the surface lowering relative to a snow/firn pack that would not experience any melting. I do not really see why this is relevant here. For me, the interesting thing to know would be how much additional refrozen mass sits in the firn column in 2018 compared to 1964, because that is a mass term that is missed by geodetic mass balance observations.

Yes, this number refers to surface lowering due to the internal refreezing (not taking into account potential mass loss due to drainage into and out of the firn aquifer). We believe that it is useful to include this number to understand the impact that refreezing has on the snowpack.

The changes in near-surface density between 1964 and 2018 that we report provide an indication of the changes in mass over time. We agree that determining the change in mass of the snowpack as a whole would be ideal, but we lack the data to know how to properly extrapolate our local data on a regional basis, so we do not do this as it would require too many assumptions and suggest a level of confidence in the accuracy of the data that is unrealistic on a regional basis.

L487-488: "to have experienced a minimum of 0.73 ±0.23 m of surface lowering due to internal refreezing": Refreezing does not lower the surface, melting does and gravitational settling of snow/firn. Please clarify and rephrase.

Apologies, rephrased.

L487-497: I am missing a bit the point here. Surface elevation changes are the effect of long-term trends in melt and accumulation. How much of the melt water refreezes does not (directly) affect surface elevation or thinning. I would rather expect a discussion here on the impact of increased densification on geodetic mass balance estimates. Geodetic mass balance observations will just consider surface height changes and not any mass changes that result from an increasing density of firn.

As stated in the reply to Line 487, we unfortunately lack the data to provide a meaningful discussion of this point in the paper.

L496: "liquid water retention processes cause the surface to lower". This is not correct. Refreezing just adds mass to the existing firn column, which leads to densification, but not to thinning! Higher firn temperatures after refreezing do speed up the compaction (gravitational settling) process though.

Thank you, we have rephrased this, Lines 607-610.

Figure 1: It could be good to include coordinate axes.

Done.

Figure 6c: In addition to Fig. 7 also Fig 6c confirms that the modelled subsurface temperatures are much colder than observed (Fig. 5).

We have removed this figure, but we do discuss the discrepancy at length in the manuscript and supplementary material.