The subject matter of this paper is highly relevant and worthy of investigation as it concerns understanding the future mass balance of glaciers, ice caps and ice sheets globally. The datasets used in this study, which include a combination of valuable field data (GPR and dGPS), and state-of-the-art firn densification, and flow models should be sufficient to address the main objectives of this study. Scientific rigor however is absent in several areas, particularly in supporting primary conclusions of this paper. For example, one of the main results, ie. PFA has persisted since the 1950's, is based on reanalysis data, for which there has been no attempt to establish quantitative relevancy to the study site. There are several instances where correlations between datasets, which are not present, would be helpful to provide more informed statements about the conclusions of this paper. Also, there has been very little attempt to explore other evidence to support the primary conclusions, as discussed below.

This paper is fraught with poor writing quality, numerous errors in punctuation, inconsistent formatting, and very poor sentence structure throughout most of the paper. Some sections however are better than others. Also, there are many problems with the figures (discussed below) which make interpretation of the results of this paper challenging. As a result, a thorough assessment of the scientific validity of this paper is not possible.

We thank the reviewer for a thorough analysis of our paper and agree that the conclusion on the PFA existence in 1957 could be made clearer. We argue that the Nordli et al (2020) timeseries are a useful proxy for the general climate of the Svalbard archipelago during the years before our modelling. We argue that this information is useful to draw conclusions on aquifer presence in 1957, as discussed in more detail in our reply to the first specific comment. We further highlight that the meteorological forcing dataset (NORA10) is elevation dependent and has previously shown good performance for temperature against AWS data at various elevations and on different glaciers in Svalbard (see Van Pelt et al. 2019 for more details). Finally, we agree that the writing and figures needed improvement and will make thorough revisions throughout the manuscript.

Specific Comments (individual scientific questions/issues)

L18 (abstract): The statement " *We find that the aquifer was present in 1957, and...*" is unsubstantiated based on the evidence provided. Specifically, the main point of evidence given within the paper for the PFA to be present in 1957, ie. "*Given that temperatures during 1957-1977 were likely cooler than during 1937-1957 (Nordli et al., 2020)...*" is not even a certainty so how can this 'finding' that the PFA existed in the 1950's be stated with absolute certainty, as it is in the abstract? Further more, the temperature data used in the model, ie. Riestad, (2011), is downscaled to ocean surface. At the very least, a correlation between the longer term Riestad data (downscaled over ocean surfaces) and the data collected from the Upsalla University AWS (~1000 m a.s.l.), should be stated.

Also, why is Nordii used as the reference for existence of the PFA, where the Riestad data was used to drive the model? Beyond the model results, this paper does not explore any further evidence to support, or question existence the PFA 'existed' in the

1950's. Is there no other evidence to support, or refute the possibility that preferential water flow routing may have been affected by changes, in ice dynamics, topography, thickness, and/or firn density over this 60+ year period of time? At the very least, it would be helpful to include a figure of near-surface air temperatures from the nearest station in order to provide the reader with some sense of the magnitude of temperature variability over the 60+ year period of this study. Simple calculations of the depth of penetration of the cold wave, and the modeled temperature regime that existed at depths to the bottom of the PFA.

We agree with the reviewer that our line of reasoning supporting the notion that the perennial firn aquifer existed before 1957 should be made clearer. The timeseries in Nordli et al. (2020) show homogenised observational timeseries for Svalbard (Longyearbyen) back to 1898. The Reistad et al. (2011) dataset only goes back to 1957. Fig. 3 in Nordli et al. (2020) reveals that the decades before 1957 (~1930-57) were warmer than the period 1957-1977. Even though these timeseries apply to sea level we argue it is unlikely that an opposite temperature trend would apply at high elevations. Since we used the colder period 1957-1977 to initialize our model, the presence of a perennial firn aquifer at the end of the spin-up period is a strong indication that a perennial firn aquifer would also have been present in our model in 1957 if we would have been able to initialize the model for the actual (warmer) period 1937-1957. We argue that higher temperatures imply more melt and growth of the firn aquifer as this is also seen during warm years throughout our simulation period. Based on the likely warmer temperatures prior to 1957 than the period that was used for model spin-up, we could rather argue that the perennial firn aquifer was larger than the size it currently has at the start of the long run in 1957. Hence, we argue that our conclusion about likely firn aquifer presence is justified, and may in fact be rather conservative. We will rewrite L296 – 298 according to the detailed comment below.

The Reistad et al. (2011) dataset was used to drive the EBFM to generate, among others, firn characteristics, as described in Van Pelt et al (2019). The NORA10 HIRLAM regional climate model based reanalysis dataset from Reistad et al. (2011) is originally available at 10-km spatial resolution and is already elevation dependent (topography is accounted for in the regional climate modelling). These data hence do not apply to sea level. In Van Pelt et al. (2019) it is described how this dataset is further downscaled to a 1-km spatial resolution and further corrections are made there to project air temperature and precipitation onto the finer-scale elevation grid. For more details on this downscaling procedure we refer to Section 2.1 in Van Pelt et al. (2019). Van Pelt et al. (2019) also compared downscaled air temperature timeseries against AWS observations on various glaciers/ice caps in Svalbard (Kongsvegen, Holtedahlfonna, Austfonna and Nordenskiöldbreen), showing good performance (see Table 3 in Van Pelt et al. 2019). For example, the nearest AWS on Nordenskiöldbreen at 520 m a.s.l. (only about 10 km's from the PFA area) shows an R-correlation of 0.95 and a bias of 0.8°C for 3-hourly temperature data between 2009 and 2015. In general, the comparison with the AWS data, covering an elevation range of 0-700 m a.s.l.) across Svalbard does not indicate that there are elevation-dependent biases in the temperature forcing dataset.

We agree with the reviewer that there could well be other factors controlling the development of the perennial firn aquifer and the way water moves through the firn. Changes in ice dynamics, topography and ice thickness are not taken into account in this study, changes in firn thickness and density are however included in our modelling. Regarding the ice dynamics, the Lomonosovfonna perennial firn aquifer is located on a very slow moving ice cap, close to the ice divide between several glaciers. The surface height change rates at our study site have been close to zero during our modelling period, see Fig. 2 in Geyman et al (2022). Hence, there are no indications that the ice geometry and flow speeds (e.g. affecting crevassing) have markedly changed throughout the simulation period.

L152 "*Furthermore, fast deep percolation is modelled using the parameterization by Marchenko et al. (2017)*". Question: Is percolation of meltwater not being modelled in 2 ways then, ie. by the Van Pelt EBFM AND the Marchenko model ?

Yes, it is indeed. 'Slow' percolation is modelled with the tipping-bucket method , fast percolation through preferential pathways is modelled by using the parameterization proposed by Marchenko et al (2017). This combined model is described by Marchenko et al. (2017) and implemented in EBFM (Van Pelt et al. 2019). We will change the text (L147 – L154) to make it more clear:

'Water in the EBFM originates from surface melt and rain, and percolates down from the surface into the firn. This happens in two ways: by using the fast deep percolation statistical parameterization from Marchenko et al. (2017), and by applying the 'tipping bucket' scheme on the remainder. The fast percolation instantaneously distributes water in the vertical model according to a normal distribution with a peak at the surface. Further water transport is modelled with a tipping-bucket approach. First, the water refreezes when the conditions in a model layer are sufficient, being that the temperature is below melting point and the density below the density of ice. Refreezing raises the temperature and density. If not all water refreezes, a small portion will be stored in the layer as irreducible water content. The remaining water will percolate down to the next layer, where the process repeats. This continues until the water encounters impermeable ice, in which case the water becomes runoff (i.e. water input in the LPFAM).'

L296-298:
 "*At the end of our spin-up routine in 1977 a PFA is present in the firn. Given that temperatures during 1957-1977 were likely cooler than during 1937-1957 (Nordli et al., 2020), we argue it is highly likely that a PFA was already present in 1957 and in the preceding decades.*"
This argument that a PFA was already present in 1957, is not convincing as it is poorly supported with scientific proof, or even a convincing argument.

We will rewrite L296 – 298 according to the answer to the first specific comment above:

' The timeseries in Nordli et al. (2020) show homogenised observational timeseries for Svalbard (Longyearbyen) back to 1898. Fig. 3 in Nordli et al. (2020) reveals that the decades before 1957 (~1930-57) were warmer than the period 1957-1977. Even though these timeseries apply to sea level we argue it is unlikely that an opposite temperature trend applies at high elevations. Since we used the colder period 1957-1977 to initialize our model, the presence of a perennial firn aquifer at the end of the spin-up period is a strong indication that a perennial firn aquifer would also have been present in our model in 1957 if we would have been able to initialize the model for the actual (warmer) period 1937-1957. We argue that higher temperatures imply more melt and growth of the firn aquifer as this is also seen during warm years throughout our simulation period. Furthermore, high melt rates were also listed as a requirement for firn aquifer formation in Kuipers Munneke et al. (2014). Based on the likely warmer temperatures prior to 1957 than the period that was used for model spin-up, we argue that the perennial firn aquifer may have been larger than the size it currently has at the start of the simulation in 1957.'

L120: this statement … '*Firn density and meltwater input to the aquifer is required to model a PFA.*' comes across as a general statement about modeling a PFA. I'm sure this is not what is intended by the author as it ignores many other factors that could be relevant. Please clarify.

That is correct. We do not want to imply that firn density and meltwater input are the only fields neccesary to model a PFA, far from it. We want to highlight here that for our approach, by using a groundwater flow model, we need as input information about the firn pack, in particular firn density and meltwater percolation. One can obtain those from observations, built a firn model within a PFA model or take those two fields as output from an existing firn model.

We will rewrite this to:
'In this model study, time-dependent firn density and meltwater input are the primary input variables used to model the PFA found at the Lomonosovfonna ice cap.'

L255-260: re: 2017 non-coherent data…. Could this not have been simulated by deriving the correlation values for the areas in common over the 3 time periods, ie remove all points from areas in 2017 that do not exist in the 2018 and 2019 datasets. This may provide some quantitative basis to refute or support the reason for coherence in the 2017 data.
Also, could there have been other factors responsible for the low coherence in the 2017 data, eg, towing speed of radar?, rough surface topography?

We thank the reviewer for this useful suggestion. We checked, and found that in 2017, only 11% of all observational points of that campaign lay within the 2018 and 2019 grid, which is about 3800 datapoints. If we would restrict our data to just the 2018 and 2019 grid, we would therefore lose almost 90% of the datapoints collected in 2017, which is undesirable.

Yes, there could have been other factors that influence the non-coherence. Missing effects from not including topography changes and ice dynamic effects can influence PFA formation and water movement. However, as stated in the answer to the first specific comment, our observational area has not experienced major geometric changes (Geyman et al. 2022) and ice flow is slow due to the proximity to an ice divide.

Furthermore, the source of non-coherence from towing speed and rough, possibly slightly different surface topography are to a first order represented by the difference in 2019 and 2018 data. Our observations show only slight differences, so we are confident in stating that the effect of differences in observational conditions (such as towing speed and slightly different topography) are small, and not enough the explain the error between modelled and observed water table depth in 2017, which is in the order of meters.

Technical Corrections

Fig 1: - Significant place names should be added to the left hand side figure. Include label location of the Svalbard airport (as it is mentioned in the text). The figure caption is poorly worded and hard to understand.

We will add Longyearbyen Airport, and rewrite the figure caption.

I don't see 2015 or 2016 on the figure, only in the caption

We removed the years 2015 and 2016 because of their sparsity in useable data points, we will remove it from the caption

- 2017 looks green on the fig, indicated as yellow in the caption

Correct, a remnant of the deleted years from the previous comments. We will change this accordingly

- In sentence 2, its not clear what 'red rectangle' you are referring to – left or right hand figure.

Left hand side, we will change the color and the sentence to make it easier to distinguish.

For this sentence...."The rectangles correspond to the minimum and maximum coordinates of the measurement in those years and thus show the extend of the PFA measurements" .

- 'Extend' should be 'extent' (2 instances), turned should be converted,

We will change this

- Avoid over-use of first-person pronouns - it is very distracting to the readability of the paper. This is particularly an issue in the results section 4.1 and conclusion.

We will reduce the amount of first-person pronouns in the paper

- Please correct citations to use semi-colon, not comma to separate a list of references. An example of this is on line 50, but it happens throughout.

We will do this

- Please follow the rules of the journal when referring to figures and tables, noting the differences between use of these words at the beginning of the sentence versus in running text. Many errors of this kind throughout the paper.

We will read the guidelines of the journal and adjust the text accordingly

- Please note that the word "Table" is never abbreviated and **should be capitalized when followed by a number (e.g. Table 4).**

We will change this

- You have 'in situ' and 'in-situ' throughout, please be consistent with formatting rules.

We will use 'in situ'

L165: remove the word 'in' from the bracketed text.

We will do this

L215: 'transfer' doesn't seem like the right word... 'transform' or 'convert' perhaps?
   -remove space before comma

We will use 'convert' and remove the space before the comma

L238:   Improper use of brackets

We will remove 'e.g.' and write the list of studies as a 'normal' end-of-the-sentence reference. Also, we will change the comma to a semicolon

L216: change 'will' to 'is'
- It is not common for a DEM to have a variable cell size - please clarify.

This is a mistake on our side, the DEM we use has a constant cell size, which is 5 meters. We interpolate it to our grid, which has rectangular cells (72 x 96 meters) to accommodate 100 cells in every direction while maintaining a rectangular area around our observation locations. We will add this to the text.

L218: the differences here needs to be stated. please indicate rmsd between the datasets

We will rewrite this, this is from an older version of the model. We do not use the topography observations anymore: we simply use the depth to the water table directly from the GPR, and use the DEM from Melvaer et al (2014) to turn LPFAM output from height above sea level to depth to the water table.

L235: close off should be close-off

We will change close off to close-off throughout the manuscript

L262: no hyphen in northwestern or southwestern

We will change this throughout the text

L263: include the max and min differences between the modelled and observed water table depths.

We will add this

L310; this sentence makes no sense on its own.

We are not sure to what sentence the reviewer is referring. The two sentences can be shortened and combined to: ' The water table depth and meltwater input at a single modelled grid cell, where the measurement station is located, is shown in Figure .... '

L325: should be Fig. not Fig

We will change this

L337: "*If refreezing happens, an ice lens might form*." This is a weird statement - Why would ice not form if refreezing 'happens'?

We will change this to 'Meltwater refreezing creates an impermeable ice layer in the firn'

L349: improve the results by how much? Enough to explain the absence of crevasses as being the reason why coherence was so much lower?

We will provide the RMSE, and add a figure to the supplementary material in which we show the location of the added crevasses, and a discussion on the additional assumptions made.

L220-240: this section is misplaced as it describes methods, not results.

We agree and will move this part to the methods section

L252: Change...

'*The modelled water table is overestimated in the south-eastern corner*'
 to
'*The modelled water table **depth** is overestimated in the south-eastern corner*'.

We will change this

L228: consequence is misspelt.

Thanks for noticing, we will change this

L320 and beyond, please refer to 'density' as 'firn density'

We will add this thourghout the manuscript.

Figure 2: y-axis – specify 'above mean sea-level'

We will change this

Figure 4. it would be far easier to compare the spatial pattern of differences between observed and modelled if the same color scheme for water table depth was used for both. Or, a third plot could be added to illustrate the differences.

We note that Figure 4 shows on the left side the modelled water table depth with the observation locations, and on the right already the difference between modelled water table depth and the observations. We see that the title of the right plots is misleading, it should be modelled water table depth difference wrt observations [year]. We will change this

It would be very useful here to have a box superimposed on the 2017 results which indicates the extent/positioning of the 2018/19 results.
Great suggestion, we will add this
- Caption refers to 'left column' and 'right'.
We will add 'right column'

Figure 5: units missing on both x and y axis.
We will add [m]
Figure 7:
- Fix caption (ie., Fig 7 to Figure 7.)
We will change this
- Color code left hand axis title to match blue line.
We will do that
- Refer to water table line as blue.
We will do this
- Not sure you need to specify both 'right' and 'orange'
We will only refer to 'orange' and 'blue'