

Dear Sebastian Gerland,

Thank you for taking the time to manage the review process of this manuscript – including your suggestions at an earlier stage, which have helped to improve the manuscript substantially. We have copied our responses to the reviewers below, including numbering of the minor comments, which we refer to in the tracked changes version of the revised paper.

Reviewer 1

This was a worthwhile project, suitable for publication in TC. I have a general comment, a major comment, and several minor comments.

Dear Reviewer, thanks a lot for your positive feedback on this study. We really appreciate your comments and have responded to them one at a time below.

General comment: Many numbers are given to higher precision than warranted. The precision of the average should not exceed the value of the standard deviation. For example, on line 246, change “ 69.5 ± 8.8 ” to “ 70 ± 9 ”. This change also makes the result easier to read. As another example, on line 295, change “ -14.6 ± 7.6 ” to “ -15 ± 8 ”.

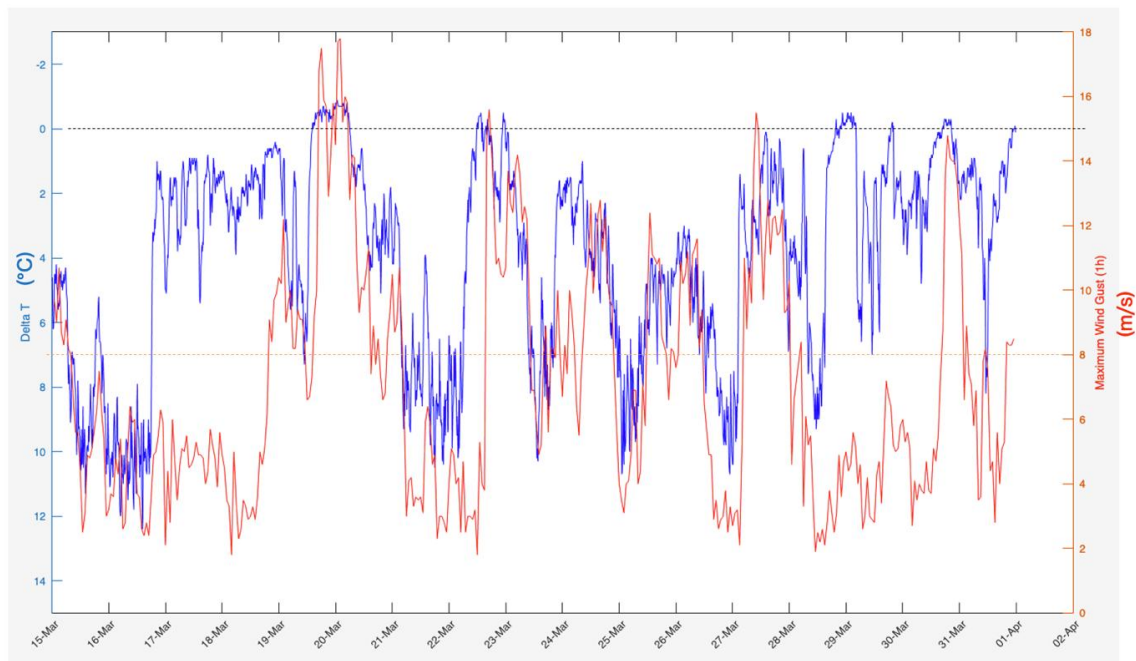
- *Thanks a lot for pointing this out. Following your advice, we have corrected the number of significant figures throughout the manuscript.*

Major comment: I am skeptical of the snow surface temperatures (SST) reported in Figure 5. The reported SST is often nearly 10 K colder than the air temperature, and is never warmer than the air temperature, although the temperature at 10-cm depth (measured by a different method) exceeds the air temperature until 30th March. How was the SST measured (maybe by IR emission)? Has the SST method been verified by in-situ thermistor measurements? [This verification should be done at night, so that solar heating of the thermistor does not cause a bias.]

- *Thank you for sharing your concerns.*

The sensor used to measure the snow surface temperature (SST) is the commercial SIR sensor manufactured by Sommer Messtechnik, which is now also reported as “snow surface temperature” in Table 1 and throughout the text. You are correct in assuming that this is a highly sensitive infrared temperature sensor designed for snow surface temperature measurements. Although a dedicated SST validation campaign has not yet been conducted and published, weekly manual measurements of the temperature within the uppermost centimetre of the snowpack, performed using thermistors, have generally agreed with the automatic observations since 2020.

For the air temperature sensor, we use a WMO-compliant and highly reliable outdoor air temperature sensor with 4-wire Pt100 1/3 DIN manufactured by NESAs. The marked temperature difference between the 2-m air and the SST (up to 10 °C on some days) could be due to several factors, including radiative forcing, cloud cover, and snow properties like snow surface area (SSA). However, we'd like to argue that one of the main drivers is wind, and particularly wind intensity. To support this interpretation, the attached figure shows the difference between T_{air} and T_{snowsurf} (Delta T. (°C); blue curve) together with the hourly maximum wind gusts measured in Ny-Ålesund during the same period (red curve).



In general, during windy conditions or periods characterized by frequent gusts (e.g., 19–20 March, 23 March, and 27–28 March, 31 March), the lower atmosphere within two metres above the snow surface is well mixed, resulting in smaller temperature differences between the two measurements at those heights. In contrast, the largest temperature differences between the snow surface and the air sensor at 2 m occur during calm wind conditions, when the atmosphere is relatively well stratified and the snow surface temperature become significantly lower than the air temperature because of radiative cooling. This behaviour is particularly evident during periods with mean wind speeds lower than a threshold (i.e. 8 m/s, considering max wind gust in NYA from SEKLIMA), such as on 16 March and around 21–22, 23, 25, and 27 March.

Although we have decided to not add this detailed exploration of the temperature distribution to the manuscript, because it's not the core focus of the study, we hope that this explanation has been reassuring regarding the data quality presented here.

To address the final concern about seemingly discrepant temperature values measured “at 10 cm depth”, we’d like to point out that the temperature measurements within the snowpack are taken 10 cm, 40 cm and 55 cm above the ground, as stated in line 208, Table 1 and the caption of Figure 5. This means that the T_{10} record in Figure 5b, which exceeds air temperature until 30th of March, refers to the lowest snow sensor just 10 cm above the ground rather than being closest to the surface. Due to the strongly insulating properties of the snowpack, this measurement is only weakly affected by daily temperature fluctuations and instead follows a more smoothed seasonal trend influenced by the underlying permafrost temperature. We explain this temperature difference between near-surface snow and deeper parts of the snow in Sect. 4.2 (line 365-366).

Minor comments:

1. Title (and elsewhere). “Constrain” is a synonym of “inhibit”. So to “constrain” the influence of near-surface melting suggests that you are looking for ways to reduce its influence. I suggest changing “constrain” to “quantify” everywhere it appears. Or say “determine”, as you do on line 394.
 - *That’s really helpful – thank you. We have changed the wording to quantify, understand and other terms both in the title of the paper and where appropriate throughout the paper.*
2. Throughout: I suggest writing dD instead of $\delta^2\text{H}$. dD is commonly used, and δ^2 looks like delta-squared.
 - *Thanks for this recommendation. We have updated the abbreviation throughout the manuscript, including Figure D1.*
3. Line 27. “increased by 4-5 C”. Since what year?
 - *Good point. We have added the reference years and, to keep the description focussed on the Svalbard region, we have revised the sentence as follows: “This is due to Svalbard mean temperatures having increased by 3–5 °C from 1971 until 2017 (Hanssen-Bauer et al., 2019) and temperatures rising especially during winter ...”*
4. Line 36. Explain why melting generally leads to ^{18}O enrichment.
 - *Thank you for this question. The enrichment arises through several complementary mechanisms.*

First, during snowmelt, isotopically lighter water preferentially drains from the snowpack, leaving the residual snow progressively enriched in heavier isotopes — the so-called melt-out effect. This is cumulative: even small-scale meltwater movement drives net enrichment, with early-season runoff consistently carrying a lighter isotopic signature than the source snow. As the melt

season progresses, both the residual snowpack and the meltwater it generates become increasingly ^{18}O -enriched.

Second, non-equilibrium fractionation during melting and refreezing — particularly during recrystallisation — tends to favour incorporation of heavier isotopes into the solid phase, further contributing to enrichment within the snowpack.

We note that these processes produce relative enrichment through internal isotopic redistribution. Absolute enrichment of the bulk snowpack additionally occurs where there is an external input, either rainfall (which, falling at warmer temperatures, carries a heavier isotopic signature than snow) or the influx of meltwater with a distinct composition.

Beria et al. (2018) have provided a good review such snow hydrological processes and their effects on stable water isotopes, so that we reference their work in the introduction of our manuscript to keep the scientific background concise.

5. Line 78. “snowpack covering 60-100% of Svalbard”. Is this a seasonal variation, i.e. 100% in winter and 60% in summer?
 - Exactly. Gallet et al. (2019) estimate: “In Svalbard, the seasonal snowpack (including glaciers) covers ~60 to 100 % of the land between winter and summer.” (Gallet et al., 2019) To clarify this seasonality, we now explain in the text: Consequently, the seasonal snowpack adding up to 40% to the ~60% spatial glacier coverage of Svalbard through the annual cycle (Gallet et al., 2019) is changing.”
6. Figure 1. Point out that all these sites are near sea-level, none on the mountain or on a glacier.
 - This is a good detail to add. We have updated the caption of Figure 1: “... (b) snow pit study sites near sea level in the vicinity of Ny-Ålesund”. We further added this detail to the following sentence: “The field campaign was initiated with an overview of avalanche-probe-based total snow thickness measurements (± 0.5 cm accuracy, provided as snow height above ground) on the local slopes near sea level on 18th March 2023.”
7. Lines 171-172. “maximum percolation depth . . . and average percolation depth”. Table E2 gives average but not maximum.
 - We have decided to focus on average percolation conditions rather than extremes reached in this study. To streamline the text body accordingly, we have removed reference to maximum percolation depth throughout the manuscript. The explanation of the measurement now reads: “Average percolation depth is the

surface melt layer thickness on the respective front and is identified through the depth reached by blue dye during an experiment (Fig. 2). “

8. Line 176. Change “percolation depth dependent” to “percolation depth to be dependent”
 - *Thanks, we updated the sentence accordingly: “We considered average percolation depth to be the dependent variable and the following independent variables: surface snow...”*
9. Table 1, Liquid water content. Give a reference for NESA-SM1, and ideally briefly say how the method works.
 - *Thanks for this comment. Liquid water content at different depths is measured using Frequency domain reflectometry (FDR) soil moisture sensors manufactured by BITLINE. Those FDR sensors installed at different depths determine moisture within a material by injecting a high-frequency electrical signal to measure the snowpack's electrical capacitance, which varies with the water content (within a sphere of 7 cm around the sensor position), and converting the resulting resonant frequency into a volumetric percentage. We mention FDR as the method that NESA-SM1 is based on in the sensor specification column of Table 1. While we consider more detailed descriptions of all the automatic nivometric station (ANS) sensors outside the scope of this manuscript, because our study focus is on the tracer percolation experiments and ANS data only briefly touched on for broad contextualisation, we reference the paper which introduced the ANS setup in more detail (Scoto et al., 2023) in the text body of Section 2.4 and caption of Table 1.*
10. Line 226. Change “temperature gradient” to “temperature difference”.
 - *We have changed the term as suggested.*
11. Figure 4c (also on Figure C1). What is the green line plotting? Its horizontal position moves back and forth slightly, but there is no scale for it on the horizontal axis.
 - *The green line in both figures displays the density profile obtained from top to bottom of each snow pit before the tracer experiments. The corresponding label and unit (kg/m^3) are provided in green font below the x axis of each panel.*
We acknowledge that the ‘skyline’ display style can be a bit challenging to read when first encountered, but (1) it was our priority to display all records for each snow pit together for full context, (2) adding another scale to the horizontal axis would probably be visually overwhelming, and (3) the display choices within the niViz software used for all snow pit panels are limited and this aspect is not manually adjustable.

To ensure that the profiles can be interpreted in absolute density terms, we have included the density values for each sampled depth interval next to the related skyline section in Figure 4c and Figure C1.

12. Figure 4 caption line 1. Change “with” to “versus”.

➤ *We have changed the wording as suggested.*

13. Line 253. The LWC is not given on Figure C1. Figure C1 includes a label “LWC”, but only gives a symbol, “D”, always the same, not a number. The symbol D is also undefined.

➤ *LWC in each snow pit was tested using the “hand test” with categories defined in Fierz et al. (2009). The term D is used for dry snow: “Usually T_s is below 0°C, but dry snow can occur at any temperature up to 0°C. Disaggregated snow grains have little tendency to adhere to each other when pressed together, as in making a snowball.” Since LWC falls in the dry (D) category throughout and for all experiment snow pits, we don’t discuss this record in the main body of the text, and it’s only provided in the appendix for completeness.*

Nevertheless, we added the reference to this Classification to the caption of Figure C1 to help readers with interpreting all material and added the definition of D to the legend of Figure C1.

14. Line 295 says “average $-14.6 \pm 7.6\text{‰}$ ”, but line 297 says “mean of $-14.55 \pm 2.5\text{‰}$ ”. What distinction is being made here? They seem like two slightly different averages for the same thing.

➤ *Thanks for your remarks on this slight difference. While $-14.6 \pm 7.6\text{‰}$ is the average ($\pm 2\sigma$) $d_{18}\text{O}$ across all measurements taken during the field study, $-14.5 \pm 2.5\text{‰}$ is the average $d_{18}\text{O}$ signature when measurements are averaged for each snow pit first and then averaged over all study sites. The consistency of both mean values and the relatively small 2σ of the second estimate are important, because this indicates that the individual snow pits analysed here are representative of the range of $d_{18}\text{O}$ signatures encountered.*

To clarify this in the text, we have revised the paragraph as follows: “SWI signatures in the seasonal snowpack near Ny-Ålesund before percolation experiments average $-15 \pm 8\text{‰}$ for $\delta^{18}\text{O}$ and $-100 \pm 60\text{‰}$ for δD across all individual measurements. The relatively large spread reflects natural variability within and between snow pits. When SWI signatures are first averaged within each snow pit and then across the study sites, the pit-level means cluster more tightly: $\delta^{18}\text{O}$ shows a mean of $-15 \pm 3\text{‰}$, ranging from -13‰ to -16‰ , and δD averages $-99 \pm 20\text{‰}$, ranging from -88‰ to -111‰ . These ranges reflect differences in SWI signatures across the Gruvebadet slope. The consistency between both mean estimates indicates that the individual snow pits sampled here are representative

of the broader $\delta^{18}\text{O}$ and δD variability encountered in the seasonal snowpack up to March 2023.”

15. Line 353 cites the extreme rainfall event of 24th March 2007. I actually was in Ny-Alesund during that event; it was dramatic. Then during the next two days the wet snow froze into a crust that the reindeer could not break through. But the reindeer knew what to do; they walked up to higher elevation where it had snowed rather than rained. According to my log for that time, the rainstorm occurred in early evening of the 23rd, not the 24th.

➤ *Thanks a lot for sharing those insights. It’s incredibly interesting and valuable to know how the local fauna copes with these events.*

Regarding the date of the event, is there any chance that the timing differed slightly across the island? Peeters et al. (2019) refers to R. Aanes for personal observations in their description of the rainfall event. We have kept the 24th of March 2007 as date for now, because, to our knowledge, there is currently no other published record mentioning the earlier timing, and the date is only of minor significance to the discussion in this study, but we’d be very happy to change it to “23rd–24th March 2007” if you could refer us to another published study or provide a name for “pers. obs.”

16. Line 421. Define “matrix flow”.

➤ *Matrix flow is first introduced including a reference for further detailed explanation earlier in this section 4.4 (line 409-411): “This suggests that there is a shift from preferential, spatially heterogeneous flow to more homogeneous matrix flow, which is dependent on the dynamic evolution and interplay of air and snow temperature. Both of these primary flow regimes have been previously documented (Moser et al., 2024, and references therein), ...”*

17. Line 421. Define Ø.

➤ *Thanks for pointing this out. We have added this definition: “fine-grained (average diameter Ø = 0.2 mm)...”*

Reviewer 2

This study has addressed the role of snowpack stratigraphy on the percolation of water through the snowpack, with the specific goal of quantifying the impact of rain-on-snow events on stable water isotope signatures which are highly important as climate proxy records in ice cores. The study carries out a well-planned field campaign to simulate the impact of ROS close to Ny Ålesund in Svalbard, with results that I consider to be significant outside of the objectives of this study too. Specifically, while the experiments simulate a minor ROS event, the results reveal that variables related to the snowpack history plays an important role in determining ROS impacts on the cryosphere – and thus ecosystems that are dependent on snow and snowmelt. Furthermore, I believe that the results also highlight why some ROS events can be more impactful than others despite similar meteorological characteristics. It would be very interesting to see results from a follow-up experiment simulating a longer and/or more intense ROS event which percolates deeper into the snowpack. I consider the manuscript to be well-written and worthy of publishing with only minor changes.

- *Dear Reviewer, thanks a lot for your positive feedback on this study. In the following, we address the detailed suggestions made.*

Minor comments

- 1) Line 77-78: “Consequently, the seasonal snowpack covering 60–100% of Svalbard”. Typically seasonal snow often refers to the non-glaciated part of Svalbard, which is ca. 40% since glaciers account for just under 60% of the total area. Perhaps consider amending this sentence to specify what you mean by seasonal snow.
 - *Thanks for pointing out this lack of precision in language. The other reviewer had a similar comment, so that we updated the sentence as follows: “Consequently, the seasonal snowpack adding up to 40% to the ~60% spatial glacier coverage of Svalbard through the annual cycle (Gallet et al., 2019) is changing.”*
- 2) L255-257: “above-freezing temperatures were reached on at least three occasions (5th – 6th, 14th – 15th, and 25th – 26th February 2023) and two day-long warm spells caused near-melting conditions (21st –24th February and 1st – 2nd March 2023).” Possibly I have misunderstood here, but were the two warm spells from 21-24th February and 1-2 March different to the occasions with “above-freezing” temperatures? Or could you just say that there were five occasions with warm/near-melting conditions?
 - *This is a good question. The overarching goal when listing the warm spells of varying degree was to point out phases that could have already changed the pre-*

existing snow structure in the area before the experiments, enabling a good description of the pre-melt snowpack characteristics in the nivometric context.

Thereby, we intended to nuance between:

- (A) events when temperature definitely reached melting point ($T > 0^{\circ}\text{C}$), and*
- (B) other events that were close but didn't reach the threshold yet (e.g. $-3^{\circ}\text{C} < T < 0^{\circ}\text{C}$).*

This seemed relevant, as we wanted to provide an overview of:

- (A) time periods when the conditions were definitely favourable to the occurrence of melting and refreezing, and*
- (B) not dismiss potential changes of the snow microstructure (i.e. grain shape and size) that look like low-grade melting/wet snow already even when air temperature hasn't widely reached melting point yet.*

As stated in Sect. 4.2 (line 360 and following), a detailed matching of melt-like horizons to specific meteorological events is beyond the scope of this study. Nevertheless, we wanted to be precise in our description of the conditions prior to the experiment campaign. To clarify the difference between the near-/melt events, we updated the sentence as follows:

"In the lead up to the fieldwork campaign during February 2023, air temperature reached values above freezing on at least three occasions ($T > 0^{\circ}\text{C}$: 5th – 6th, 14th – 15th, and 25th – 26th February 2023), and two sub-zero yet warm periods ($-3^{\circ}\text{C} < T < 0^{\circ}\text{C}$: 21st – 24th February and 1st – 2nd March 2023) could have also enabled changes to the snow structure."

- 3) L395 onwards. The discussion of the MLRA results is interesting, in particular how difference in grain size at stratigraphic boundaries can constrain the vertical flow but it was not obvious to me whether there were cases used in the analysis where the tracer had percolated deep enough to reach ice layers from previous ROS events/warm spells, and if so how did it interact with the ice layers eg. Does the water pool and refreeze at existing ice layers (thereby thickening them) or do they melt out the ice layers?? I think it was already mentioned elsewhere in the literature that it can be challenging to disentangle the signatures of multiple ROS events in the snowpack and I think this is an important result too since the number of ice layers in the snowpack will not necessarily correlate with the number of ROS events.
- *We absolutely agree with your comments and that observations of melt–ice layer interactions are urgently needed in glacier hydrology research. Unfortunately,*

none of our experiment tracers reached pre-existing melt layers, so that we can't comment on their permeability and related behaviours here.

Regarding the other aspect of difficulties in disentangling multiple ROS events, we already briefly touch on it in Sect 4.3 to the degree that it is evident from our study simulating individual ROS events. "The spatial heterogeneities of melt lenses and PFPs on a centimetre-to-decimetre scale and internal layering on a millimetre scale (Fig. 6) also demonstrate how challenging it is to structurally differentiate individual melt events or melt-intensive seasons in firn and ice cores (e.g., Alley and Anandakrishnan, 1995; Moser et al., 2024)." We have added the reference to Alley and Anandakrishnan (1995) and Moser et al. (2024) to highlight that there is a wider literature context for this discussion and added one sentence to point out more clearly: *"Moreover, the number of ice layers in the snowpack will not necessarily be equal to the number of ROS events that have occurred during the season (e.g., Pfeffer and Humphrey, 1998)."*