We would like to thank the referee for their detailed and very useful review. We appreciate the reviewer's time and effort. We believe the manuscript is now in a much better shape. We hope this manuscript help the community in future approaches for SWE retrieval missions. We had two main changes to respond properly to the reviewer's concern.

- We used the line with zero interception for fitting.
- We used the average of two stations for calibrating the retrieved SWE.

We would like to thanks the reviewer for pointing to these two problems. The first one was needed and the second one make it easier for the readers to follow.

Please find our responses below.

RC1, Anonymous Referee #1

Review, Oveisgharan et al. "Snow Water Equivalent Retrieval over Idaho, Part A: Using Sentinel-1 Repeat-Pass Interferometry"

<u>General comment:</u>

The manuscript performs an analysis of a 6-day repeat pass InSAR time series with the aim to retrieve snow water equivalent (SWE) from the InSAR phase. The theoretical foundation of this SWE-retrieval method is known for more than 20 years but common InSAR problems like coherence loss, phase ambiguities, and the choice of the best reference value to correct for unknown atmospheric phase delays still pose significant obstacles for the method, in addition to the limitation to dry snow. The analysis of a time series with one of the shortest currently (locally) available repeat intervals of 6-days at a relatively low frequency (C-band) is a valuable contribution to assess the best combination of repeat-pass interval and frequency.

However, in its current form the manuscript has several shortcomings that needs to be addressed before consideration for publication:

1) The linear approximation of the equation from Guneriussen (2001) has a methodological shortcoming: a linear approximation (y = mx + c) is performed and after the approximation the y-axis intersection c is neglected. Correct would be to first assume the y-axis intersection is zero, and then estimate the slope m. Even though the estimated slope changes by several percent, it would not significantly impact the results of the paper; however the authors claim to provide a more general approximation compared to existing literature which is simply not correct; furthermore, the comparison of their estimates with literature (Fig. 10a) shows a significant deviation due to the methodological fitting mistake. In addition, their approximation method is not reproducible due to the lack of given ranges for snow density and incidence angles (see specific comments on section 2.2 and the attached figure).

- You are right. This is one of the two sections we changed significantly. We used c=0 and then did the fitting.

2) In section 5.1 the authors use the same data (average of all station's in-situ SWE measurements) for calibration of the InSAR based SWE estimate and for validation (local station in-situ SWE vs. local InSAR based SWE estimates). This means that the calibration data is not independent from the validation data (see specific comments on line 242-248/Figure 7a and b, 172a/b).

- I do not agree completely. We changed the average of all stations to average of 2 reliable stations and the results degrades a little. Even though we used the average of all stations to match the average of retrieved values, it's not similar to saying that we shift each station separately. This is something that has been done for soil moisture to remove the bias. The more the number of stations are the more the average is dependent to one or two of them or biased towards one of them. However, we understand that this will may confuse people and so we used two stations for calibration and the rest of them for validation.

3) In section 1 and 2, several statements are supported by reference to literature, however, in a very vague, not correct way or misleading way. These references should be improved (see specific comments below).

- We addressed all the ones you mentioned in the revised version of the paper.

4) In my opinion, low correlation coefficients (0.4...0.6) are often overrated by the authors as "very high/highly correlated/one of the best metrics". They need a more critical consideration.

- We changed the high correlated to correlated. However, w believe 0.5 is relatively good correlation. You mentioned 0.9 is a good correlation. I agree with that but the two datasets should be almost the same to have a correlation of 0.9. For now, I tried to tone it down.

5) Noteworthy, in section 5.2 the authors provide a reasonable correlation (0.51 and 0.62) between their spatially distributed SWE estimate and completely independent LiDAR data which supports the feasibility of SWE retrieval from 6-day repeat pass C-band data.

- Thanks! Yes, this was the highlight of this work.

<u>Specific comments:</u>

Abstract:

4: "optimal": Before specifying what the optimal parameters are, I would not use the work "optimal". I suggest "suitable"

    - Done!

6: "long time series": In the sentence before, you talk about repeat intervals, therefore please specify if your time series has a 6-day or 12-day (or any more complex combination) repeat interval.

- We changed the sentence to: "In this study, we apply this technique to a long time series of 6-day temporal repeat Sentinel-1 data from the 2020-2021 winter."

10: "highly correlated with LIDAR": The correlation with the SWE station data (r = 0.82) is much stronger than with LIDAR data (r = 0.5). Therefore I would not say "highly correlated" which indicates that also a very good relation between snow height and SWE exists. However, dependent on the snow pack, density variations can deteriorate SWE estimations from snow height. Writing "highly correlated" here implies that both, radar interferometry and LIDAR estimates are equally precise which is physically not correct, because radar interferometry shows a physics-based, almost linear relation to SWE while LIDAR estimates require a guess (or auxiliary information) of the average snow density (as written in line 29). I suggest something like "well correlated".

- That's fair. We changed the highly correlated to well correlated.

32-38: I suggest a more critical review of the listed papers. The references are given in a context that suggests that SWE estimation with active sensors is generally possible from space. I think that is misleading and contradicts the abstract (line 3). Did all authors use spaceborne sensors and did they really estimate SWE? I suggest to better detail and quantify where the authors of the listed papers have been successful and where not - i.e., where are still current gaps that have not been filled by the listed authors?

- We rewrote this paragraph trying to explain the limitation of each. The limitations for the ones with the interferometry technique are explained later in next section. "Active microwave sensors provide high resolution and global coverage. There has been many efforts in the last two decades trying to estimate SWE or snow depth using active sensors mounted on a tower (Cui et al., 2016; Lemmetyinen et al., 2018; 35 Ruiz et al., 2022; Leinss et al., 2015), airborne (Marshall et al., 2021; Nagler et al., 2022), or spaceborne systems (Lievens et al., 2019; Liu et al., 2017; Conde et al., 2019; Dagurova et al., 2020; Eppler et al., 2022). Backscattered power from active sensors is used to estimate SWE (Rott et al., 2010; Ulaby and Stiles, 1980; Cui et al., 2016; Nghiem and Tsai, 2001; Lievens et al., 2019). A dual-band (X and Ku) SAR mission has been the focus of the European Space Agency (ESA) and Canadian Space Agency (CSA) for SWE spaceborne measurements (Rott et al., 2010; Lemmetyinen et al., 2018). However, accurate *a priori* characterization of snow micro-structural parameters is of primary importance in the accuracy of SWE retrieval algorithms using backscattered power (Lemmetyinen et al., 2018; Durand and Liu, 2012; Cui et al., 2016). The most common *a priori* characterization used for SWE retrieval algorithms using backscattered power is grain radius. This has been done using passive data; however, the methods are restricted to passive retrieval errors and also mismatch between active and passive resolutions. The ratio of cross-pol to co-pol Sentinel-1 backscattered power has been used to estimate snow depth over mountainous regions with deep snow (Lievens et al., 2019, 2022). Using Sentinel-1 backscattered power ratio is a unique approach showing the success of snow depth retrieval using the spaceborne radar time series data. However, the retrieval mostly works for deep snow in mountainous regions. The radiative transfer physics at C-band for this method are still poorly understood.

47-49: These lines imply that FMCW radars cannot be used in space because of their wide bandwidth. That is not correct. FMCW radars are cheaper because they do not require temporal pulse compression (that needs high-power electronics) but because they transmit and receive at the same time, they are limited to a range of a few km because the signal travel time limits the pulse repetition frequency. And it would even be simpler to build them with a narrow bandwidth, but a benefit of these experimental and locally installed systems is the possibility to explore wide bandwidths. It is more, that wide-band system cannot be used in space because of limited frequency allocation. Possibly, I misunderstood your point: Did you mean nadir-looking radar altimeters (that could be FMCW from air or pulse-systems from space). Please clarify.

- FMCW systems needs to have wide bandwidth in space to avoid ambiguity. Something that we do with prf for pulsed signal. We should be able to estimate the distance from the received frequency. With that wide bandwidth for large distances we have the problem of bandwidth allocation for active sensors. We add this clarification to the paper: "These sensors need a wide bandwidth for spaceborne large distances. Due to allowable frequency bandwidth of a spaceborne active sensor (Tao et al 2019), FMCW system cannot be used in spaceborne missions for global coverage due to their wide bandwidth."

50: "specularly reflected": that is not correct in the context of the given references. "specularly reflected" means "like on a mirror", according to the law of reflection. However, the listed authors observe the snow/ground surface from the same point as the illumination source (monostatic or almost monostatic radar system).

- The authors wanted to show the theory behind SoOp (specular reflected signal) is very similar to differential interferometry. But we realized it hasn't been communicated correctly in the paper. Now it is rewritten: "The phase change of specularly reflected signals in Signals of Opportunity (SoOp) is shown to be strongly dependent on SWE changes for dry snow and on depth changes for wet snow (Yueh et al., 2017, 2021; Shah et al., 2017). The phase sensitivity to SWE changes increases at higher frequencies. However, the temporal coherence decreases and the phase ambiguity increases which makes the phase unwrapping very challenging at higher frequencies. The theory behind this method is similar to repeat pass interferometry that is explained in section 2. The advantage of this method is that the stratigraphy of the snow has little impact on the SWE retrieval (Yueh et al., 2017) similar to SWE retrieval explained in section 2. Using the long wavelength signal at P-band in SoOp is very helpful for addressing the loss of temporal coherence and phase unwrapping challenges of this method. However, there has been very limited data showing the success of this method at P-band. Achieving high resolution is another challenge of this method."

52: "in wet snow the phase center is normally at the snow surface": that might be true for X-band and above, but the cited papers used P-band where I would expect a significant penetration into the wet snow pack. Did the cited authors really show that the phase center is at the surface? Please clarify or correct.

- It has been shown in section III.E of Yueh et al., 2017, that snow wetness brings the reflection from the ground-snow to air-snow. They also show that the phase change is proportional to snow depth for wet snow. You are right that the signal should penetrate to some extent into the snow at P-band but the main signal will come from the snow ground interface for specular reflection. For backscattered differential interferometry geometry (our retrieval method in this paper) the phase may not be that linearly dependent to snow depth change.

53: "the phase unwrapping (...) increases at higher frequency." That does not make sense. Either "the effort for successful phase unwrapping increases at higher frequencies" or "the phase wraps more frequently at higher frequencies."

- That is correct, thanks for pointing that out. We changed it to: “The phase sensitivity to SWE changes increases at higher frequencies. However, the temporal coherence decreases and the phase ambiguity increases which makes the phase unwrapping very challenging at higher frequencies.”

55: "the theory behind this method": Could you first concisely describe the method you are referring to? What's the core idea of the method (see Guneriussen 2001)? After that, you can outline successful applications and limitations of the method as attempted in line 50-54.

- We think with the new write up of this paragraph (written in response to 50), it is clearer.

56: snow stratigraphy and grain size: you are citing Yueh(2017) here, but consider also citing Leinss(2015) who did an error analysis on the impact on SWE estimates from layers with different density.

- Good point, that one is more complete, thanks for pointing that. Done!

80: Please provide a reference after "dual-pol dual-frequency retrieval algorithm". Are you sure that this algorithm relies on the assumption of non-scattering, non-absorbing dry snow, so that microwaves penetrate to the ground?

- The reference is added. We also add a sentence to better explain why we need the wave to penetrate all the way to the ground. We need to make sure that the phase change is due to phase delay from the ground not phase center shifted to somewhere in the snow volume. Here is our modification “Similar to the dual-pol. dual-freq. retrieval algorithm (Lemmetyinen et al., 2018; Cui et al., 2016), this technique relies on the dryness of snow in order to penetrate all the way to the ground and the scattering from the snow layers and snow volume is minimized compared to snow-ground return.

90: "highly correlated": Please quantify by providing a correlation coefficient.

- Done by adding it to the text: “The correlation of 0.76 was observed between the retrieved SWE change using L-band UAVSAR differential interferometry between 2/1/2020 and 2/13/2020 and the collected LIDAR snow depth change between 2/1/2020 and 2/12/2020 over the open regions of Grand Mesa in dry snow conditions (Marshall et al., 2021).”

95: "works well": Could briefly you point out how you solved the problem of phase reference and how you corrected for atmosphere?

- In section 4 we explained the details of our retrieval. Here we are just explaining how our approach is different from others as our retrieval covers a long time series with many in situ stations for validation. Describing the detail of reference point and atmospheric removal, even briefly, seems out of context here. However, for your reference, we used mintpy to remove atmospheric noise and we used the average of reliable in situ stations as the reference point.

106: "a high temporal coherence is observed...with a month temporal baseline": Could you provide some quantification or statistics from the cited paper? Was this "high coherence" observed in a single observation or was it generally high over a certain time series? Under which snow conditions?

- We changed it to “A medium mean temporal coherence of 0.41 is observed at L-band between two winter seasons in shrub-lands with 10.2cm average snow depth (Molan et al., 2018)” to make it clearer. The paper is not focused on snow in particular but has used the interferograms over snow and snow is one parameter in their modeling.
- Figure 2(c) of the cited paper shows coherence between 0.4 to 1 for one month observation in shrubland but because the snow depth change isn’t mentioned we decided to not use that and talk about the average between two winter season.

109: What do you mean with "controlled system"? What snow conditions did the authors observe?

- We removed the controlled system as we thought it may be confusing for the reader. What we meant was that they were able to distinguish the vegetation was almost frozen in the snow and doesn’t move, they distinguish one or two occasions that there was windy nights, … . However, in a real big scene we cannot distinguish these scenarios. In addition they mentioned “All acquisitions were coregistered to compensate range-shifts due to a snow related signal delay” which needs some information from the scene for that I believe.

110: "the ... temporal coherence... Leinss (2014)": The cited paper does not deal with temporal coherence. Please check. I guess, you mean Leinss(2015).

- That is correct. We fixed it.

114: "the temperature was shown to be the most critical variable ...": In which sense? There are strong differences between positive and negative temperatures. Variations in negative temperatures should not significantly affect the coherence.

- We added this sentence to the paper “Temperature above zero reduced the temporal coherence drastically”. Although as the temperature decreases below zero the coherence increases gradually in most cases and different frequencies in the referred paper.

120: I would drop the complete sentence "More studies are needed..." because it is very general.

- Agreed and Deleted!

Section 2.2:

Comment for section 2.2: The described method seems to contain a few methodological shortcomings, misses values of parameters that are needed for reproducibility and is, in my opinion, less general than the approximation of Leinss (2015). Even though I do not expect any major impact on the results, I would like to urge the authors to improve this section and to set it in proper context with existing literature, specifically Gunneriussen (2001) and Leinss (2015) but also Mätzler (1996) and Wiesmann (1999) [reference below, comment on l.128]. An interesting idea of the described method is to fit a function for every incidence angle - this idea is, however, almost identical to the introduction of the fit parameter \alpha in Leinss (2015) to provide an optimal numeric approximation for every chosen incidence angle, because the incidence angle is very often precisely known - in contrast to the snow density, for which only a reasonable range can be estimated. In contrast to Leinss (2015), where a density range from 0 to a maximum snow density (rho_max) was assumed, a slightly improved approximation could possibly be achieved by assuming/selecting/limiting snow density to a more realistic range, e.g. rho = 0.1...0.5 g/cm^3, or whatever the authors expect for their specific region.

- As you can see below, we changed the section to fit a line to it. Thanks for pointing this out. However, we still think using one equation for the entire incidence angle range is very convenient. So, by this analysis we can see what the error is using this approximation. However, we can use different fitting for each angle too. However, it makes things much faster for a big scene to use just one formula.

Specific comments for section 2.2:

125: "related the interferometric phase directly": Could you be more specific? I suggest "related the interferometric phase with a linear approximation directly to SWE changes"

- We changed it to “With some approximation to equation 1, Leinss et al. showed a linear relationship between the interferometric phase and SWE change (Leinss et al., 2015).”

126: "depends on the range of incidence angles and snow density": That is not exactly correct. I suggest: "depends on the chosen incidence angles and the maximum expected snow density"

- Fair enough, we changed it to: “The approximation depends on the incidence angle of each pixel and the maximum expected snow density.”

126: Could you specify what exactly you mean with "more generalized"? In my opinion, your method is less general then the approximation by Leinss (2015), because it is limited to snow densities up to 0.5 g/cm^3 and incidence angles between 20 and 50 degree.

- I think the fact that we have “one” formula for any point in the image to convert the phase to SWE makes it generalized. You are right that we know the incidence angle and we use the corresponding formula for that incidence angle but I think it makes it more time consuming. The density is limited to reasonable snow densities observed and incidence angle is what we generally observe in Sentinel-1 images. So, in applying the formula to a SAR image in Snow it is generalized. Although the incidence angle can get larger as can be seen in figure 1(c) and the error remains less than 10%.

127: What are the units that you use for density? kg/m^3 or g/cm^3 or volume fraction (unitless)?

- It is g/cm^3 as explained in the figure but added here too.

128: "We use Matzler's model for calculating epsilon in equation 1 (Mätzler, 1987).": Could you provide a more specific reference (e.g. equation number or show the used equation for epsilon)? The cited dissertation has 130 pages and contains many different models for the permittivity of snow. Note, that there are more recent equations for the permittivity of snow from Mätzler, e.g.:

- We added the corresponding equations in the text.

  - C. Mätzler, "Microwave permittivity of dry snow," IEEE Trans. Geosci. Remote Sens., vol. 34, no. 2, pp. 573–581, 1996-03, doi: https://doi.org/10.1109/36.485133.

  - A. Wiesmann and C. Mätzler, "Microwave Emission Model of Layered Snowpacks," Remote Sens. Environ., vol. 70, no. 3, pp. 307–316, 1999, doi: http://dx.doi.org/10.1016/S0034-4257(99)00046-2 (Note, that Eq. 46 misses an exponent of 1/3 for eps_s which becomes apparent when comparing with Eq. 10, Sect. IV-F in Mätzler 1996).

129: Please specify what units you use for SWE: kg/m^2, mm w.e.q., m^3/m^2=m (w.e.q)? Make sure, that the equation returns the correct values when entering SWE, rho, k_i, and C in the specified units.

- We double checked all the unites and add a sentence to be clearer. "Note that the ε and consequently C are unitless, $\rho_{water} = 1$ g/cm^3, ΔSWE$ is in (m), and κ_i is in (1/m)."

Figure 1a: What units has C(theta, rho)? Could you verify if you really plotted C(theta, rho) as defined in Eq. (2). When I compute C from equation 2, I obtain the data shown in Fig. 1a, but only when not dividing by snow density rho. That means, currently, Fig. 1a is identical to Figure 8left, in Leinss (2015), and C(theta, rho) == xi(theta, rho_s); an overlay of both figures confirms this. I think the division by rho in the definition of C should be removed.

- You are right, thanks for noticing that! I didn't notice this is the same as figure 8 in Leinss paper. I moved the rho to equation. However, in terms of simplification of equations and plots, the rest are correct.

Note:

For the rest of the comments for this section, I completely agree with you. Thanks for your great suggestion! I think I was fitting and then notice B is very close to zero, … . But as you mentioned it makes more sense to assume B=0 from the beginning. I rewrote this section assuming B=0. I also specified the ranges for incidence angle and density. I agree that depending on that range the "A" changes. Although it doesn't change the results significantly. So, I use incidence angle range of Sentinel-1 data over this frame (0 to 80) and snow densities as mentioned before. Using the new approximation I redid all retrievals in the results section. The results for in situ stations remain almost the same and the correlation in lidar section degrades very little as expected.

132: "we fit a line to C for different incidence angles": Could you specify the interval of densities rho that was used for fitting? Depending on the chosen interval, the fitting parameters can vary by several percent (see attached figure).

- Rewrote the section.

132: Could you specify the interval of theta for that lines-slopes were obtained by fitting? In the attached figure I assumed 20..50°, similar to Figure 1b and 1c.

- Rewrote the section.

135: I cannot reproduce the equation in line 135. When following the described approach (assuming a density range of rho = 0.15..0.5 g/cm^3) my values for the parameter A deviate by 1-2% from eq. 135. (see attached figure, dashed red line vs. solid blue line). Also note that for a different density interval of rho = 0.1..0.4, the values of A deviate 3-4% from eq. 135 (dashed purple line).

- Rewrote the section.

136: "Assuming B(theta) = 0": I see this as the main shortcoming in this section: Why do you first fit a line including an y-axis intersection, and then assume that the y-axis intersection is zero? As shown in the attached figure, you obtain more accurate results when first assuming "B(theta) = 0" and then fit a line C = A(theta)*rho where the y-axis intersection is zero by definition (solid teal line). The difference to eq. 135 is about 5% at incidence angles of 20°.

- Rewrote the section.

136: "B is very close to zero": (See also comment above). How close? Neglecting terms makes only sense when putting them into relation to larger terms. Could you put B in relation to A * rho? The ratio of B/A is up to 1.5%, depending on the chosen fitting interval; however, you need to consider the equation C = A(theta)*rho + B(theta) where the snow density (assuming units of g(cm^3)) can be very small, again, depending on the chosen interval. With rho = 0.15, B is as large as 16% of (A*rho) which can have - and has - a significant impact on the fitting results.

- Rewrote the section.

139: What's the advantage of using eq. (4) vs. Eq. (18) from Leinss 2015? Considering the above described shortcomings, I do not see any advantage of using this equation. But a proper linear fit as suggested two comments above would do it.

- I think our equation is applicable to wider range of snow densities and incidence angles which makes it convenient for Sentinel-1 frame.

Section 3.1:

143-151a: This paragraph reads than a general description about Sentinel-1 and the ASF. Could you describe which data and which processing workflow you use for your input data?

- We edited to better show what we used for our workflow: “We used the Interferometric Wide Swath (IW) mode data with 5 and 20m single look resolution in range and azimuth direction, respectively. The IW swath width is about 250km. We used ASF On Demand Processing to generate interferometric phase and coherence at vv and vh (transmit-received polarization) polarization. Alaska Satellite Facility's Hybrid Pluggable Processing Pipeline (HyP3) is a service for processing Synthetic Aperture Radar (SAR) imagery. The workflow includes interferometric phase correction for ground topography and geolocation. The ASF HYP3 uses a Minimum Cost Flow (MCF) algorithm for phase unwrapping. The unwrapped phase and interferometric correlation were used in this study.”

143-151b: Which incidence angle (range) did the acquisitions have that you processed?

- The incidence angle for the frame we processed varies between 0 to 80.

152-172: same as above: I would expect an introducing sentence to motivate why you discuss the SnowEx2021 campaign(s) and the SNOTEL data. The reason why these campaigns are discussed is given at the end of the section, line 171-172. It would be easier to read if these sentence appear as a motivation/reason at the beginning of the paragraph. (see also next comment).

- I added this sentence right before the SnowEx campaign as an introduction: “Sentinel-1 collects data every 12 days globally but has the capability to collect the data every 6 days over targeted areas, mainly over Europe and selected areas such as SnowEx sites. In order to validate our SWE retrieval using Sentinel-1 data, we use LIDAR data from SnowEx campaign and SNOTEL data as discussed in section 5. We also use the average of SNOTEL data as a reference point for SWE retrieval, as seen in section 4.”

162/Figure 3b: Could you add to the caption in Figure 3b that for Delta SWE the first date of the differenced dates is shown?

- Done. We add this sentence to 3(b):” Note that the Delta SWE is marked on the first day of each observation.

163: "located in remote, high elevated mountainous regions": Figure 3a shows strongly variable SWE values. Is that due to altitude differences of the stations? Could you specify the altitude range of the stations?

- This is nice feature of using Spaceborne data. It covers a wide range of stations with different SWE values. And using the average of SWE change of all those stations doesn’t necessarily bias the result. We add this sentence to the text:” Different colors show different SNOTEL stations. The elevation of these stations varies between 3200m to 9520m. Therefore, the large spread of SWE between different stations in figure 3(a) is expected.”

170: "As seen in this figure [3a], there is a one 6-day repeat data acquisition gap in Sentinel-1 data on 2/5/21": Figure 3a shows two gaps as if there would be no acquisition on 2021-01-30 which is not true (Figure 4/5 show the pair 2021-01-24 vs. 2021-01-30).

- As mentioned in the paper, the dashed lines show the first day of each interferograms. There was observation on 01/30 but hasn’t been shown here as there is no 6 days observation that starts on 01/30. We understand that it is a little confusing but we want to identify the dates as appear in figure 3(b).

172a: "we used these in situ data for SWE retrievals performance": what do you mean? To tweak the performance? Or to validate the performance?

- Changed it to: ”SWE retrieval validation”.

172b: [used for] "the InSAR reference points": Could you clarify which data you used for InSAR phase reference points and which data for validation of InSAR-based SWE results?

- We add this sentence for more clarity: “We used the SWE data from these in situ stations for (a) SWE retrievals validation by comparing retrieved ΔSWE with SNOTEL ΔSWE(as seen in section 5.1), and (b) the InSAR reference point by subtracting the average of two SNOTEL ΔSWE from the retrieved ΔSWE (as explained in section 4).”

174-175: "LIDAR...are reliable sources of validation data, particularly a powerful constraint for InSAR retrieval of SWE": same as comment above: Could you clarify if you used the data for validation or to constrain the results?

- We added the sentence:” We used the LIDAR data for validating the retrieved SWE results.”

175: "The "SnowEx20-21 QSI LIDAR DEM 0.5m" data set is part of the SnowEx 2020 and SnowEx 2021 campaigns (Adebisi et al., 2022). The data includes digital elevation models, snow depth, and vegetation height with 5m spatial resolution." - Could you clarify which data had 0.5m resolution and which 5.0 meters?

- Thanks for noticing that, it was a typo. We changed the 5m to 0.5m.

185: Could you explain why you selected the period from 2020-12-01 to 2021-03-30? Did you limit the period to dry snow conditions?

- We added this sentence to clarify: “We selected this period to (a) capture most of the seasonal snow storm and (b) avoid wet snow as much as possible.”

185-186: Do you have any information about (or did any correction for) the height-(or air pressure) dependent phase contribution? It seems yes, maybe add a reference to Section 4.1.

- I am not sure what you mean here. We use pyapps to remove troposphere noise which is dependent on the topography, and we mentioned that reference in section 4.1. So, I am not sure what you are requesting here.

189/191: "In this study, any pixel with less than 0.35 temporal coherence is not considered reliable (...) However, in .. 5.1.2 and 5.2 we used all ... data even with low coherence to calculate total SWE": Why do you consider it as an advantage to use all data, even though you consider the data as unreliable due to low coherence? Or is there something specific about these two sections that the reader has not yet been informed yet? Would it not be more of advantage to set unreliable data to some assumed SWE change, e.g. zero? (see also comment on line 197)

- We changed it for total SWE calculation and now we just use the data that is reliable. However for the LIDAR data, we used all the data because we are calculating total SWE and if we miss on day i it will affect the total SWE on day i+n. So, if we have a lot of SWE on day i, we are ΔSWE_i less in the coming days, hence divergence. So, we consider all the data assuming there is some error in that. We saw that even with that assumption, the results are still good enough. If we consider zero for SWE on day i, our results are ΔSWE_i away from the actual total SWE at the end. We added this sentence to make it clearer: "The reason is that in order to compare the total SWE at each day, we need the whole ΔSWE time series up to that date."

192: "The ionospheric error ... is much smaller than other sources of error .. and considered negligible": Could you distinguish tropospheric error and ionospheric error? Or is the ionospheric error removed together with tropospheric phase delays?

- Ionospheric error is due to ionosphere delay and is negligible at C-band. It mostly lies between 85km to 600km above earth surface. The atmosphere is ionoized and contains plasma in this region. The total electron content is what affects the phase delay of InSAR. It is dispersive for interferometric phase which means that it is proportional to square of wavelength. Therefore, the effect at L-band is much bigger. We ignore its effect on C-band Sentinel-1 data. Troposphere is mostly up to 12km above earth and is nondispersive meaning that it doesn't depend on the radio frequency wavelength. We remove it by models as explained in section 4.1. We added this sentence for more clarification: "The radar signal propagating through ionosphere is delayed. The delay is a function of frequency of the signal, Earth's magnetic field, and total electron content (TEC) and affects the accuracy of ΔSWE retrieval. The ionospheric error at C-band is much smaller than other sources of error and we consider it negligible in this study."

197: "similar to correlation filtering, for ... 5.1.2 and 5.2 we used all the ... data": why? (see comment 189/191)

- We changed it for section 5.1.2 and just compared the total for reliable SWE but still kept all the data for LIDAR comparison. We understand that it will affect our accuracy but there is no other workaround solution. We add this sentence: "Similar to correlation filtering, for the results in section 5.2, we used all the time series data, even with temperature more than zero. Similar to temporal coherence, the reason is that in order to compare the total SWE with LIDAR snow depth at LIDAR acquisition date, we need the entire ΔSWE time series up to that date."

201-203: For choosing the phase reference point, Tarricone et al. (2023) point out a promising idea for mountainous terrain: They suggest snow free areas, derived from optical data (fractional snow cover maps), as phase reference. I think it's worth to cite this paper here.

- Tarricone, J., Webb, R. W., Marshall, H.-P., Nolin, A. W., and Meyer, F. J.: Estimating snow accumulation and ablation with L-band interferometric synthetic aperture radar (InSAR), The Cryosphere, 17, 1997–2019, https://doi.org/10.5194/tc-17-1997-2023, 2023.

- Thanks for pointing that out. We add it to the paper: “For ΔSWE estimation using InSAR, the reference point is chosen either by corner reflectors (cleaned of snow) with stable zero phase (Nagler et al., 2022; Dagurova et al., 2020) or using the average of in situ ΔSWE (Conde et al., 2019) or using a snow free region (Tarricone et al., 2023)”

Note:

We changed the reference point from average of all SWE values to average of two in situ stations with good temporal coherence and temperature less than zero for the entire time series. This way we avoid the same calibration and validation problem. However, for correlation coefficient calculation we need to remove the mean of the two variables, by definition. So, I think even using the average of all stations shouldn’t matter much. Note that our calibration doesn’t involve any model fitting. It is just the subtraction of a number from the entire image. The subtraction of mean of in situ station is being used for soil moisture unbiased estimation. In any case, I think using the average of just two stations would address most of your comments for this section. The results degrade a little bit but not significantly. Below please find the responses to each one.

204-206. "we used the average of all in situ Delta SWE ... to calibrate the retrieved [InSAR] Delta SWE images": Figure 4 (and also Figure 3) shows strongly variable SWE values across all stations. Why would a correction with the average improve the estimates? Would a correction using stations with small/no SWE change (I would assume, these are the stations at lower altitude) not be more beneficial (see also Tarricone TC 2023)?

- Normally the places with low snow change and as you mentioned low altitude go through melting more regularly. The melting and temperature are one of the main sources of error in this method. Even for the areas with no snow change soil moisture affects the phase and consequently SWE change(you can see that in phase closure studies in recent years). On the other hand, in the areas with lots of SWE change, phase unwrapping and phase ambiguity are a challenge with low temporal coherence. So, average of all may not be the best option but it is one way to go. However, as mentioned we used two stable in situ stations now for calibration and this is not a problem anymore.

210-211: Could you provide the given information (Delta SWE, and the information that there was no SWE change and two storms) (also) in the caption of Figure 4?

- Done! We add this sentence to the caption: “The average of in situ ΔSWE, for images (a), (b), and (c) are 0.01cm, 2.72 cm, and 4.33 cm, respectively.”

242/244: "all the retrieved Delta SWE" / "all the retrieved Delta SWE time series": It seems you plotted Delta SWE between each two S1 acquisition for all SNOTEL stations. Even though you analyzed data from the whole time series of images, the shown data does not contain any specific time-series characteristics (in contrast to section 5.1.2). Is that correct? If so, clearly indicate this.

- That is correct. We add this sentence for more clarity: “Note that the data shown in figure 6 is the SWE *change* between two consecutive Sentinel-1 data that are 6 days apart. We showed the ΔSWE for all stations and all consecutive observations between 12/1/20 and 3/30/21.”

242-248/Figure 7a and b: What is the correlation coefficient between the mean Delta SWE values of all stations (that was used as reference for the Delta SWE images - see line 204-206) vs. the Delta SWE of all individual stations? - The contrast of the good correlation between 0.6 and 1.0 (with a mean around 0.82) in Figure 7b (and the mean in Fig. 7a) and the low correlation (between at least -0.5 and +0.8, median correlation at approximately 0.4) is a hint that the major part of the correlations is due to the fact that validation data was used for calibration.

For estimating a lower limit for the expected error in your data: how large is the standard deviation for certainly snow free areas (if there are any)?

- We changed the calibration method and used just two stations. So, hopefully this resolves the issue here. I didn't get what correlation coefficient you are after. As mentioned above, snow free regions may bring soil moisture information in SWE retrieval. So, that cannot be a good choice for calibration nor for minimum SWE error calculation.

252: I would not call an correlation coefficient of 0.4 and higher "very good" (See examples on https://en.wikipedia.org/wiki/Pearson_correlation_coefficient). I would call a correlation of 0.8 good, and a correlation of 0.4 poor.

- We changed very good to good. 0.8 is considered highly correlated. 0.4 you can definitely see there is a relationship.

Figure 7c: What is the correlation of the first acquisition at 12/01? Could you scale the y-axis so that all datapoints are shown?

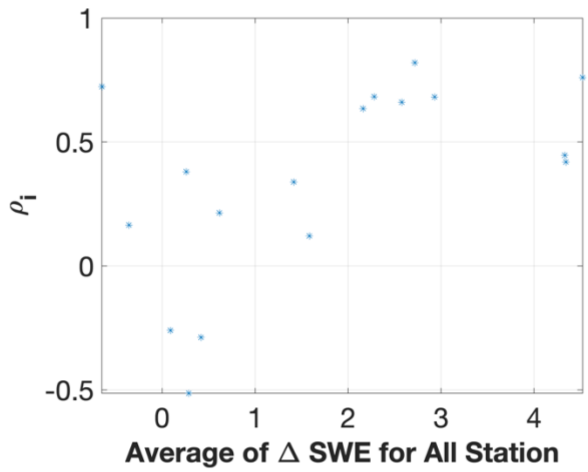
- The SWE change between day 1 and 2 is zero. Therefore, the correlation coefficient is NaN.

Figure 7c: Could you add to the caption in Figure 7c that for correlation and the first date of the correlated date pair is shown?

- We added this sentence to the caption: "Note that the labels on x-axis show the first date of each interferometric observation."

250-262: Here, you discuss reasons why you could expect a low correlation. If you select only acquisitions pairs with SWE changes (average over all station) larger than a certain threshold (according to Figure 3a) how would the correlation coefficient look like? For example, you could have the average SWE change per acquisition pair (sorted by magnitude) on the x-axis and the correlation coefficient of all SNOTEL stations vs. InSAR Delta SWE on the y-axis, would the correlation show an increasing trend with increasing SWE change? I would expect that if there's a clear correlation between Delta SWE and the station data. In contrast to Figure 7a, such a plot would be independent on the choice of the reference phase (or reference Delta SWE).

- The argument that when ΔSWE is small our retrieval doesn't have enough accuracy and therefore we won't expect to get a good correlation between retrieved and in situ is different from saying that ΔSWE is correlated to the correlation between in situ and retrieved. Having said that I plotted as you suggested for my curiosity and as you can see below there is some dependency. However, there are just 18 data points, and we cannot conclude anything. In any case, I think my point here is that for days that we have very small SWE change, it is expected that we don’t see good correlation between in situ and retrieved SWE as the noise is bigger than SWE change.



Section 5.1.2: Note, that the methodology in this section might suffer from the same calibration problem as commented on 242-248. (validation data used for calibration). If you plot the SWE error at the end of the season (03/31) over total SWE, I would expect that for stations with a large total SWE the S1 data show a negative bias, while for stations with a small total SWE, S1 shows a positive bias.

- I believe with the new reference point currently defined; this issue is resolved.

277/281: "we think the main reason for divergence is the ... phase ambiguity" vs. "The divergence is mainly due to phase ambiguity": These two statements are not compatible with each other. Do you assume/think that that's the main reason? - then you can't make the second statement. Or do you have sufficient evidence that that's the reason? - than provide the evidence it, instead of "thinking" it's the reason.

- We changed the “ the divergence is mainly due to phase ambiguity” to “We will investigate the reason behind the divergence of retrieved SWE from in situ SWE of these stations in the future work of this study.”

Figure 10a: I think the main reason for the discrepancy between the approximation from Oveisgharan (2023) and Leinss (2015) is the shortcoming of correct approximation of the equation from Guneriussen (2001). See attached figure (blue line vs. thin black line; even more accurate: blue line vs. thick black line). See comment on line 136.

- You were right. As mentioned above, I used a line going through zero and the result is very close. Thanks for noticing that.

304-309 vs. 314-318: These two paragraphs show a very high redundancy of text. Consider writing it more concise, e.g. first a single paragraph describing both figures (11 and 12), then a second paragraph describing the differences.

- Done as requested, good suggestion, thanks!

308-309: "The high correlation of 0.51 shows the success of this method": I would say "there's likely some correlation supporting that SWE estimation with InSAR from space is feasible."

- The patterns are very similar in the image and that amount correlation is not small. We changed the writing to: “The relatively high correlation (0.47 and 0.59) between the two independent measurements with different resolutions is a very good indication of the success of this method in estimating SWE.”

318-319 "The correlation ...is 0.62. This high correlation is one of the best validation metric for SWE retrieval using the InSAR techniques". First, I would not consider a correlation of 0.62 has high. It basically shows that there is likely some correlation. Considering that the comparison of the spatial snow height distribution from LIDAR is a completely independent dataset from the spatial distribution of the InSAR estimate (that includes calibration by in-situ data from SNOTEL station) makes this correlation, more correctly, the two correlations shown in Figure 11c and 12c, currently the most convincing correlations in this manuscript.

- We agree that this result is the most convincing result we showed in this work. Having said that the other ones are quite impressive. As mentioned for the previous comment we changed the writing to better address your concern: “The relatively high correlation (0.47 and 0.59) between the two independent measurements with different resolutions is a very good indication of the success of this method in estimating SWE.”

326: "highly correlated (0.82) with in situ values" - as commented above, there might be a calibration problem as commented on 242-248. (validation data used for calibration).

- With the new calibration method, that problem doesn’t exist anymore.

327: "The retrieved total SWE has less than 2 cm RMSE compared with in situ values in 16 stations" - you neglect here, that it is worse for the other 34 stations of the in total 50 stations.

- Not all of those stations are in our image. There are 43 stations in the rectangle shown in figure 4 but 12 of them are in the blue region where there is no data there. Therefore, we have 31 stations inside Sentinel-1 frame. I changes figure 6(b) to only show the 31 stations and not the ones that are out of frame. As added now to the paper: “Among all 31 stations in the Sentinel-1 frame, 6 of them have temporal coherence less than 0.35 or temperature more than zero in their entire time series. Two of them are used for calibration of the phase. So, there were 23 stations with more than 2 reliable observation dates in their time series. Among the 23 stations, 9 have SWE error less than 2cm (green diamonds) and 14 of them have SWE error larger than 2cm (red diamonds).” So it is actually not that bad.

328-329: "We show ... that SWE retrieval using spaceborne InSAR timeseries is a very promising candidate for future SWE missions": It might be true that the analysis of InSAR time series is a promising candidate for future SWE missions. However, your paper shows that the method is likely feasible, but it also shows that several problems of the method, like phase unwrapping, loss of coherence, choice of reference phase, are still not sufficiently solved and that even with a 6 day repeat interval at C-band it is not easy to retrieve reliable SWE estimates.

- That is true and we also mentioned that in the next paragraph. “We also showed that the main constraints for this method are its temporal coherence, phase unwrapping, and phase ambiguity. We showed that snow storms reduce the temporal coherence significantly. Low temporal coherence reduces the accuracy of the interferometric phase and unwrapping algorithm. Small SWE ambiguity at C-band makes the phase unwrapping more challenging. Going from C-band to lower frequencies such as L-band improves both the temporal coherence and SWE ambiguity. With the L-band NASA-ISRO SAR mission (NISAR) launch coming next winter, the new dataset would be a great dataset for global SWE retrieval.” I added “we showed” to this paragraph to confirm that we showed these problems in this study too.

I'm looking very much forward to the first SWE time series from NISAR :)

- Yes, let’s go NISAR!

Technical comments:

Please check the paper against the submission guidelines https://www.the-cryosphere.net/submission.html, specifically dates should be written as "dd month yyyy" (https://www.the-cryosphere.net/submission.html#math) rather than in American notation. I know that sometimes (e.g. in figures or tables) the date can be given according to the international ISO 8601 standard (https://en.wikipedia.org/wiki/ISO_8601) but that needs to be confirmed with the editors.

189: You might want to change double negation to a positive formulation "less than .. not considered reliable" -> "higher than ... considered reliable"

- Changed it to :” In this study, any pixel with temporal coherence more than 0.35 is considered reliable.”

Figure 4 and 5: It might make sense to combine both figures into a six-panel image with a top row of the three Delta SWE images and a bottom row with the three correlation images.

- Done!