We thank the reviewer for the careful reading of the manuscript and for the constructive comments and suggestions, which have helped improve the clarity and quality of this work.

RC3, Anonymous Referee #3

This study evaluates Sentinel-1 repeat-pass InSAR retrieval of snow water equivalent (SWE) using comparisons with airborne LIDAR snow depth and SNOTEL observations across several SnowEx sites. The results suggest that 6-day repeat acquisitions can recover spatial SWE patterns reasonably well at some sites, while performance degrades substantially for 12-day revisits. The discussion on temporal coherence, temperature, and SWE variability is also potentially valuable for future operational applications.

However, I have several concerns regarding the novelty, the clarity of the methodology, and the interpretation of the validation results. In particular, I am not fully convinced that the manuscript sufficiently distinguishes its contribution from the authors' previous work, and several of the methodological assumptions require stronger justification. I therefore recommend that the paper undergo major revisions before it can be considered for publication.

1. I am not sure the manuscript presents sufficient methodological novelty relative to the authors' earlier work. The SWE retrieval framework, phase-to-ΔSWE conversion, and even the reference-point calibration strategy appear to follow Oveisgharan et al. (2024) quite closely. As it stands, it is not very clear what is fundamentally new here beyond applying the same framework to more sites and comparing with LIDAR and SNOTEL datasets. Is the novelty primarily the broader evaluation across multiple SnowEx sites? Is it the comparison against LIDAR snow depth and SNOTEL? Or is there a methodological improvement that is not sufficiently highlighted? At present, the paper reads more like an application and extension of an existing framework than a clearly new methodological contribution. The authors need to state much more clearly what is fundamentally new in this paper and why this new step is important enough to warrant a separate publication.

- Very fair point, I address this in several places in the paper to make it clear the main contribution of the paper
    - In the abstract we added: "While previous work demonstrated the feasibility of SWE retrieval using Sentinel-1 interferometry over limited sites, this study provides a systematic, multi-site evaluation across diverse snow and land-cover conditions to identify the key factors controlling retrieval performance."
    - At the end of introduction "While (Oveisgharan et al., 2024) established the feasibility of SWE retrieval using Sentinel-1 interferometry and demonstrated its performance over a limited number of sites, this study extends that work by providing a systematic, multi-site evaluation across diverse snow and land-cover conditions. In particular, we investigate the environmental and geophysical factors controlling retrieval performance, including temporal coherence, temperature, vegetation, and

terrain, and validate the approach using a significantly larger set of LIDAR and in situ observations. This analysis provides new insight into where and under what conditions InSAR-based SWE retrieval is reliable, moving beyond proof-of-concept toward broader applicability. The methodology developed here is also directly applicable to the recently launched L- and S-band NASA–ISRO's NISAR mission."

- At the end of section 2 (before 2.1): "The main contributions of this study are: (1) a comprehensive multi-site validation of InSAR-based SWE retrieval using an expanded set of LIDAR and in situ observations, (2) a systematic analysis of environmental and geophysical factors controlling retrieval performance, and (3) identification of the conditions under which reliable SWE retrieval is achievable using C-band SAR data."
- Beginning of section 4: "Unlike (Oveisgharan et al., 2024), which focused on demonstrating the feasibility of SWE retrieval using Sentinel-1 interferometry over a limited number of sites, this section evaluates retrieval performance across a broader range of snow conditions and land-cover types to identify the key factors controlling retrieval accuracy."
- Beginning of section 5: "While previous work (Oveisgharan et al., 2024) established the capability of InSAR to retrieve SWE, it did not systematically assess the environmental and geophysical controls on retrieval performance. Here, we extend that work by analyzing how factors such as temporal coherence, temperature, vegetation, and terrain influence retrieval accuracy across multiple sites."
- At the beginning of section 5.1: "This analysis builds on the limited LIDAR validation presented in (Oveisgharan et al., 2024) by incorporating a larger number of LIDAR scenes across diverse environments, enabling a more comprehensive assessment of retrieval performance and its controlling factors."
- At the beginning of section 5.2: "In contrast to the site-specific validation in (Oveisgharan et al., 2024), this section leverages multi-site SNOTEL time series to investigate how environmental variability affects SWE retrieval performance and temporal coherence."
- End of the section 5: "This study moves beyond demonstrating the feasibility of InSAR-based SWE retrieval by providing a systematic evaluation of its performance across varying environmental conditions. The results identify the dominant controls on retrieval accuracy and define the conditions under which the method is reliable, which are critical steps toward operational SWE monitoring using current and future SAR missions."
- At the beginning of the conclusion section: "While previous work demonstrated the feasibility of SWE retrieval using Sentinel-1 interferometry, this study provides the first comprehensive, multi-site evaluation of retrieval performance and its controlling factors. By quantifying the roles of temporal coherence, temperature, vegetation, and terrain, we establish the conditions under which InSAR-based SWE retrieval is reliable, representing a key step toward large-scale and operational applications. These results also provide a foundation for quality assurance (QA) in future SWE products by identifying the key parameters that control retrieval reliability and uncertainty."

2. I do not quite follow the reconstruction of total SWE from interferometric SWE change. Equation (1) estimates ΔSWE between two acquisitions, but Equation (2) appears to reconstruct total SWE by summing these increments from the first Sentinel-1 date after 1 December, while

assuming zero SWE at that initial date. This assumption needs more explanation. In some of these mountain regions, there may already be substantial early-season snow accumulation by early December. If so, the retrieval would miss a non-negligible baseline SWE and the reconstructed total SWE would be systematically biased low. The manuscript currently treats this as a simplifying assumption, but I think it could directly affect the interpretation of the comparison with LIDAR snow depth. How sensitive are the reported correlations to this assumed zero baseline? Does this matter only for absolute magnitude, or could it also affect the spatial comparisons if early-season accumulation is not spatially uniform? I would like to see either a sensitivity analysis or at least a more rigorous discussion of the implications of this assumption.

- We added this to address this concern: "The assumption of zero SWE at t1 may introduce a bias in the reconstructed total SWE in regions where early-season accumulation is present. However, this assumption primarily affects the absolute magnitude of SWE rather than its spatial variability. For example, based on in situ measurements distributed across the Sentinel-1 frame (p71, f444), the spatial standard deviation of SWE is approximately 5 cm at the beginning of December and increases to about 23 cm by the end of March. This indicates that early-season SWE exhibits relatively low spatial variability compared to peak winter conditions. Given that the LIDAR scenes cover a much smaller spatial extent (~ 16 km × 15 km) within the larger Sentinel-1 frame (~240 km × 240 km), the spatial variability of SWE at the beginning of December is expected to be even smaller at the LIDAR scale. Therefore, neglecting the initial SWE mainly introduces an approximately constant offset in the reconstructed SWE and is not expected to significantly affect the spatial correlation with LIDAR snow depth. However, the reconstructed SWE values may be systematically biased low in absolute terms."

3. I am not fully convinced by the comparison between retrieved SWE and LIDAR snow depth without a clearer discussion of snow density. The manuscript reports correlations between Sentinel-1-derived SWE and LIDAR snow depth and interprets these as retrieval skill. However, SWE and snow depth are not equivalent quantities,4. 5. 6. and the relationship depends on snow density, which can vary substantially in space and time. It would help if the authors clarify whether this comparison is intended only to assess relative spatial patterns rather than physical equivalence. At present, this point is underexplained, yet it underpins one of the main validation results of the paper

- We dedicated section 4.1.4, "Evaluating the Retrieved Density", to address the discussion of density derived from comparing SWE and snow depth. Figure 9 addresses the reviewer concern.

4. The uncertainty analysis is still too limited. The manuscript mentions several sources of uncertainty, including atmospheric phase delay correction, phase ambiguity, wet-snow effects, and low temporal coherence. However, these are discussed somewhat separately, and I think the paper would be much stronger with a more systematic uncertainty assessment. For example, how

large are the expected errors associated with residual atmospheric correction over complex terrain? How often might the dry-snow assumption be violated even when near-surface air temperature remains below 0∘C? How much of the retrieval degradation during larger storms is due to phase ambiguity versus decorrelation? At present, the paper concludes that 12-day repeat data perform poorly and 6-day repeat data perform better, but the contribution of each error source to that difference is not quantified clearly enough. This needs to be discussed in a more integrated way.

- We added 2 sections, (4.1.5) and (4.2.1), on uncertainty analysis for LIDAR and SNOTEL comparisons.

5. I am not fully convinced by some of the broader interpretations regarding the dominant controls on retrieval performance. The manuscript argues that temporal coherence is the dominant factor controlling retrieval quality, with additional roles for temperature, SWE variability, vegetation, topography, and other environmental parameters. While this is plausible, some of these conclusions seem largely descriptive and based on a limited number of sites or frames. There are also exceptions discussed in the text, such as RC20, TU21, and the weaker consistency between the LIDAR-based and SNOTEL-based analyses. The manuscript excludes TU21 from the later parameter-impact analysis because of low overall coherence, and RC20 is explained post hoc by frequent melt events and low SWE. So I am not sure the manuscript yet provides enough evidence to conclude that temporal coherence is the dominant control in a general sense, rather than one important control among several interacting factors.

- We soften the claim and acknowledge interactions and variability in abstract: "Analysis of retrieval drivers indicates that temporal coherence is a primary control on performance, with additional contributions from temperature, snow wetness, and vegetation cover. The influence of these parameters on temporal coherence is only partially consistent with their effect on SWE retrieval performance, highlighting the complex interplay among environmental and observational factors. Temporal coherence generally declines with increasing snow depth, slope, and temperature, but improves under dry, cold conditions and gentle terrain."
- We also added this to section 5.1 to show that we acknowledge the exception, and we cannot generalize the temporal coherence impact: "At the same time, the presence of exceptions indicates that temporal coherence alone does not fully explain retrieval performance, and that additional environmental and observational factors must also be considered. Therefore, temporal coherence is used here as a primary metric for evaluating retrieval quality, while its interaction with other controlling parameters is examined in Section 5.1.1.

6. I do not quite follow the discrepancy between the LIDAR-based and SNOTEL-based performance analysis. In the LIDAR comparison, the message is that higher temporal coherence generally leads to better performance, and this is used to motivate the parameter analysis in

Section 5.1.1. However, in the SNOTEL-based analysis, the interpretation appears less straightforward, with performance being discussed in relation to temperature, SWE magnitude, and coherence in a less consistent way. The manuscript also uses SNOTEL measurements both to establish reference points and to evaluate retrieval behavior, which makes me wonder whether there is some circularity or at least a dependence between calibration and evaluation that should be discussed more explicitly. I think the authors need to explain much more clearly why the LIDAR-based and SNOTEL-based analyses do not lead to a cleaner and more unified picture if temporal coherence is indeed the dominant retrieval control. At the moment, the two validation approaches seem to support overlapping but not fully consistent interpretations.

- Fair point. We expand our discussion in section 5.2 (marked red). We hope this address your valid concern. “This behavior differs from the more consistent positive relationship observed between temporal coherence and performance in the LIDAR-based analysis (Figure 13(a)), where higher coherence generally leads to improved agreement. It is important to distinguish between the two validation approaches used in this study, as they assess different aspects of retrieval performance. The SNOTEL-based analysis evaluates SWE change (ΔSW E) between individual interferometric acquisitions, and therefore reflects the temporal performance of the retrieval at specific time intervals. In contrast, the LIDAR-based analysis compares total retrieved SWE with total LIDAR-derived snow depth at a given date, which effectively integrates SWE changes over time and represents an aggregate measure of performance. As a result, the two approaches differ in their sensitivity to errors. In the LIDAR comparison, total SWE is obtained by summing ΔSWE over time. Any error in the reference point used for phase calibration propagates as a constant offset in the reconstructed total SWE. Such an offset does not affect spatial correlation with LIDAR snow depth, which primarily reflects the similarity of spatial patterns. In contrast, the SNOTEL-based analysis compares ΔSW E across multiple dates, and is therefore more sensitive to reference point errors and temporal noise. For example, estimating a small ΔSW E for large ΔSW E can significantly affect correlation with SNOTEL measurements, whereas in the LIDAR comparison the same error would contribute only a constant bias with no on spatial correlation. More broadly, the LIDAR-based analysis evaluates "spatial" correlation of "total SWE", while the SNOTEL-based analysis evaluates "spatio-temporal" correlation of "SWE change". Consequently, the relationship between temporal coherence and retrieval performance differs between the two approaches. In the LIDAR comparison, higher average temporal coherence over the time series generally leads to improved performance, as it reflects more reliable phase measurements accumulated over time. In the SNOTEL analysis, however, the relationship is more complex. High temporal coherence often occurs during periods of small ΔSW E, when retrieval performance is inherently more sensitive to noise and may degrade despite favorable coherence conditions. Therefore, the LIDAR-based analysis captures the integrated, average behavior of the retrieval, while the SNOTEL-based analysis reveals the instantaneous performance and its dependence on temporal variability. Together, these complementary perspectives provide a more complete understanding of the factors controlling SWE retrieval performance. If a temporally consistent LIDAR time series were available, co-registered with the InSAR acquisitions, the comparison would become more analogous to the SNOTEL-based analysis, as both would involve evaluating SWE and snow depth changes between acquisition dates and their dependence on temporal coherence. In that case, the relationship between coherence and retrieval performance would be expected to follow similar behavior to that observed in the SNOTEL comparison."

7. The manuscript separates the LIDAR analysis into 6-day and 12-day revisit groups and states that retrieval performance is very poor for the 12-day repeat case. It also states that temporal coherence is generally low at C-band and that the 6-day repeat significantly improves coherence compared to the nominal 12-day repeat. However, I am still not clear what the main limiting factor is for the 12-day case. Is it mainly decorrelation over longer repeat intervals? Is it because larger accumulated. ΔSWE between visits makes phase ambiguity more likely? Is it because longer intervals make wet-snow episodes and transient melt more likely to contaminate the signal? Or is it all of these together?

- We expanded this in section 4.1.3 (marked by red). We hope it addresses the reviewer's concern. "The results (not shown here) indicate that temporal coherence degrades significantly at the 12-day revisit interval, rendering the data unreliable for SWE retrieval. This degradation arises from a combination of factors rather than a single dominant limitation. First, the longer temporal baseline increases decorrelation due to snowpack evolution, wind redistribution, and melt–freeze processes, which reduces the reliability of the interferometric phase. As discussed in Section 4.1.5, low temporal coherence leads to a substantial increase in SWE retrieval uncertainty and degrades the performance of phase unwrapping algorithms. Second, the magnitude of SWE change between acquisitions is significantly larger for 12-day intervals compared to 6-day intervals (e.g., Figure 8 versus Figure 4). For Sentinel-1, the SWE ambiguity corresponding to a 2π phase cycle ranges from approximately 1.5 to 3.5 cm depending on incidence angle. In many cases, the accumulated ΔSW E over 12 days exceeds this ambiguity threshold, making phase unwrapping more challenging and increasing the likelihood of phase errors. While such ambiguities can sometimes be resolved using high-coherence reference areas combined with in situ constraints, the simultaneous presence of low coherence limits the effectiveness of these approaches. Finally, longer revisit intervals increase the probability of transient melt events and wet-snow conditions occurring between acquisitions, which violate the dry-snow assumption required for the interferometric SWE retrieval. In addition, the reduced number of acquisitions in a 12-day time series increases the relative weight of each interferogram in reconstructing total SWE, making the retrieval more sensitive to individual errors. Overall, the combination of increased temporal decorrelation, larger SWE changes relative to phase ambiguity, reduced phase unwrapping reliability, and a higher likelihood of wet-snow conditions makes SWE retrieval from 12-day Sentinel-1 data significantly more challenging. It is worth noting that upcoming missions such as NISAR, despite having a similar nominal 12-day repeat cycle, operate at longer wavelengths (L- and S-band), which are expected to maintain higher temporal coherence and provide larger SWE ambiguity thresholds (approximately 6–14 cm). These characteristics are anticipated to mitigate some of the limitations observed for C-band Sentinel-1 data."

Minor comments:

1. Line 265: "Retrieavl" should be "Retrieval"

- Done!