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.
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.
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,

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

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.
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.
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.

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?

Minor comments:

1. Line 265: “Retrieavl” should be “Retrieval”