Title: Mapping Seasonal Snow Melting in Karakoram Using SAR and Topographic Data
Manuscript Number: egusphere-2024-942
Journal: The Cryosphere
Date: February 7, 2025

Comments from reviewers / Revisions to manuscript / **Responses from authors** (black)

General Comments

Thank you so much for these thoughtful comments, revisions and additional plots, I really appreciate the effort and thoroughness you put into your response. I feel the motivations and clarity of the manuscript are much improved and sound great! My only remaining concern is our discussion on TSI and its pixelwise multiplication with WSI.

My original concern here was that the introduction of TSI and its pixelwise multiplication with WSI was biasing your wet snow detection–TSI is essentially an aggregate backscatter response since it is the median WSI of all of its member pixels. By multiplying a given pixel's WSI by the aggregate response of the bin to which it is a member (TSI), you are strongly forcing that pixel towards the classification of the majority of pixels in their topographic bin, greatly reducing the impact of its individual backscatter response.

I appreciated your response to this, pointing out that you had used the bin median of WSI (instead of mean) to calculate TSI, and that your threshold choice of 3.5 over 5 moderated the influence of TSI. I also wanted to thank you for going out of your way to make those requested plots with classification metrics aggregated by topographic bin in Figures 1-3 (response doc), they're really helpful to see.

I think you're definitely right that using the bin median over bin mean is preferable here, but I don't think my concern is primarily about the skewness of WSI in a bin nor TSI robustness, it is more about TSI representing the aggregate behavior of that bin and how its pixelwise multiplication with WSI can overwhelm individual responses of pixels in that bin.

Regarding your choice of the 3.5 coefficient, after examining your sensitivity analysis in Figure 4 (response doc) and its relationship to the distributions shown in Figure 5 (response doc), I am a bit skeptical about whether the coefficient choice of 3.5 moderates TSI's influence in the way suggested. The TSI histogram shows values strongly concentrated in the range of 2-4 in Hunza and Shyok, with Shigar mostly 0-2. Flipping back to the sensitivity analysis, it seems the empirical coefficient is related to each basin's distribution of TSI. Given that TSI was designed on a 0-10 scale (due to how WSI was designed), perhaps these optimal coefficients are primarily compensating for this concentration of TSI values, basically "de-scaling" SI by dividing out typical TSI values, in order to align SI thresholds with the traditional -2dB threshold? In this way, the coefficient optimization seems to be addressing the introduction of TSI and the mathematical artifact of the multiplicative combination of WSI and TSI rather than representing a relationship between basin topography and snow. If you believe this is not the case, I think that the need for and purpose of this coefficient should be motivated and clarified further.

The provided classification metrics reveal how some of these TSI concerns manifest in systematic variations in precision and recall that are seemingly highly influenced by TSI. For the Hunza Basin, you show in Figure 4 (manuscript) during summer the very distinct and sudden TSI differences between areas above/below 5000m. Compare this plot to Figure 1-3 (response doc) where you show F1, recall, and precision respectively.

For the relevant elevations 2000-6000m, notice how your TSI plot and your F1, recall, and especially precision have some very similar patterns. The shape of the TSI >∼ 1 blob is approximately the area where F1 >∼ 0.9. The recall plot also seems heavily influenced by TSI, implying that you don't miss much wet snow between 5000-6000m, but you miss a lot of wet snow below these elevations. Precision tells a similar story, with fewer false positive identifications of wet snow at the lower elevations, but a lot of false positives between 5000-6000m. This area seems particularly troubling, with a region of seemingly near constant precision of 0.5 across diverse terrain conditions–maybe I'm misinterpreting this, but is this plot saying for areas covered by your large TSI >∼ 1 blob, when you guess that there is wet snow, you are only correct half of the time?

While I trust that your F1 scores and confusion matrices do show an improvement over the Rc method, I can't help but feel that these improvements may be partially arising from enforcing expected topographic patterns. This approach seems to act as a smoothing function, basically saying that where the wet snow line roughly is (from TSI), give all pixels higher in elevation a much greater chance of being identified as wet snow, and all pixels lower in elevation a much smaller chance of being identified as wet snow. This makes me think that your method's primary increase in accuracy comes from smoothing out areas with unexpected wet/dry snow occurrences–in the summer example, wet snow at lower elevations or dry snow / no snow at higher elevations.

This is why I wondered earlier about whether you performed the full Nagler et al., 2016 method, which includes resampling Rc to 100m and a post-processing step which applies a 3x3 median filter to remove outlier pixels. If your method acts as a sort of low pass filter, I'm wondering how your method performs relative to the full Nagler et al., 2016 method which already includes some simple filtering?

Sorry for the wall of text - to distill this down a bit into a short list of addressable remaining suggestions:

1. Consider adding a sentence or two with some explicit physical interpretation of TSI and the WSI/TSI/SI multiplication step, especially regarding how the coefficient used to calculate the SI threshold relates to TSI's scale and distributio
   - Perhaps could be added around lines 192 or 200-201?
2. Consider adding another sentence or two more thoroughly describing where and what type of errors are expected, especially as it relates to precision and recall
   - Maybe right after the new sentence "Finally, the use of TSI introduces potential bias in regions where topographic conditions deviate significantly from the assumptions underlying its calculation."
   - I think you should include the provided plots of F1, recall, and precision as a figure in the supplement–I think you mentioned this in the response document, but I didn't see it reflected in the tracked changes document
3. Consider comparing your method with the full Nagler et al., 2016 method (including the resampling / filtering steps) for a more earnest comparison and to better understand if the improvements in accuracy your method demonstrates have anything to do with smoothing

I apologize if this is asking for more of your time, you've already done a great job of being responsive to feedback. If time and scope are a concern, I think suggestions 1 and 2 should be quick, and for suggestion 3 maybe you can just mention as potential future work? The core contributions of your paper remain valuable regardless, and I appreciate your thoughtful engagement with these methodological questions. Thank you for your hard work, it's been a pleasure getting to review yall's paper!

Response:

We sincerely thank the reviewer for the detailed and thoughtful feedback. Your extensive comments have provided us with valuable insights that have helped us improve the clarity and robustness of our manuscript. We appreciate your recognition of our improvements so far and your careful consideration of the methodological aspects of our work. We have addressed the listed suggestions as below.

1. Physical interpretation of TSI and the WSI/TSI/SI multiplication step

We have added the following sentences to line 192 to further clarity the physical interpretation of TSI:

> *Example TSI distribution for Hunza at different elevation, slope and aspect are presented in Figure 4.* In this example, TSI values have shown different patterns across seasons and topographic conditions. In spring, strong TSI signals are found around 4000 m.a.s.l for east facing slopes (aspect between 0 $\sim$ 190°) over flat terrain (slope $\theta < 20°$), while no obvious TSI signal is observed for steep terrain (slope $\theta > 20°$). This can be explained by the limited snowmelt during the spring season of Karakoram. In summer, strong TSI signals are observed above ∼5000 m.a.s.l. for all slope and aspect conditions, indicating the presence of wet snow. However, steep slopes showed an unevenly distributed TSI across slope aspects, which can be attributed to the shadowing effect of the surrounding terrain. In autumn, TSI signals generally decrease due to the absence of wet snow, and the snow line retreats to higher elevations. This example indicates that the dynamic influence of topography on snowmelt can be effectively captured by the designed TSI signal.

We have also added the following sentences to lines 200-201 to clarify the relationship between the coefficient used to calculate the SI threshold and TSI's scale and distribution:

> *In contrast, the TSI is time-varying and basin-specific, requiring an optimal coefficient to condition the SI for classification.* This coefficient should moderate the influence of TSI in the SI formulation and compensate for the concentration of TSI values within a limited range, so that the impact of TSI can be normalized to align with traditional -2 dB reference. *The choice of this coefficient...*

2. Where and what type of errors are expected

We added the following sentences at line 301 to describe where and what type of errors are expected:

> *Finally, the use of TSI introduces potential bias in regions where topographic conditions deviate significantly from the assumptions underlying its calculation.* As shown in Figure 7, the long-tailed distribution of TSI values reflects the cumulative statistical nature of TSI, which relies on using a single median value to represent each topographic bin. This approach may be inadequate in bins with strong terrain variations, especially when the TSI distribution of pixels within a bin is highly skewed. In such cases, using the median can lead to systematic over- or under-estimation—overestimating in left-skewed distributions and underestimating in right-skewed ones—ultimately affecting precision and recall in the classification results (see the supplementary materials for detailed visualizations).

We have also included the supplement as a separate file in our manuscript submission, following the submission guidelines of the journal.

3. Comparing the method with the full Nagler et al., 2016 method (including the resampling and filtering steps)

We have added the following sentences to the end of section 5.1 as suggested:

> It is also important to note that while we followed Nagler's method (Nagler et al., 2016) to generate the R_C image, we did not apply the same post-processing steps, such as median filtering and land cover masking. These smoothing and filtering steps may influence accuracy, and future work could incorporate them to further evaluate their impact and potentially improve snow mapping performance.

References

Nagler, T., Rott, H., Ripper, E., Bippus, G., and Hetzenecker, M.: Advancements for Snowmelt Monitoring by Means of Sentinel-1 SAR, Remote Sensing, 8, 348, https://doi.org/10.3390/rs8040348, 2016.