Response to RC2

Mapping Seasonal Snow Melting in Karakoram Using SAR and Topographic Data

Discussion: https://doi.org/10.5194/egusphere-2024-942

Comments from the reviewers are given in black.

Our responses are given in blue.

General Comments

This paper proposes an improvement to the Nagler family of SAR-based wet snow mapping methods, making use of a gaussian mixture model and newly proposed metrics “wet snow index”, “topographic snow index” and “integrated snow index” in order to better separate wet snow from no snow / dry snow by incorporating terrain data and a more dynamic clustering algorithm. This method is evaluated against Sentinel-2 derived snow cover maps, and the authors found improved wet snow identification performance when compared to the Nagler et al., 2016 method. The authors apply their method over the Karakoram, a challenging domain for wet snow identification, to analyze snowmelt dynamics from 2017 to 2021.

I appreciate the authors focus on improving SAR-based wet snow mapping–there is still a lot of room for our community to improve these methods, and this research does a nice job focusing on two important considerations of these algorithms: the incorporation of terrain information and more complex clustering methods. In this paper, I particularly enjoyed the beautiful plots (especially figures 5 and 8!) and the application of this method for time series analysis over multiple melt seasons. This was really great to see, thanks for yall’s hard work!

However, I believe the authors should address some important methodological concerns and should also have a chance to improve the overall clarity of the paper. In particular, I think the newly proposed metrics need to be given more motivation and context, especially when it comes to the physical interpretation of different mathematical and methodological operations. I am particularly concerned with the topographic snow index and the implication of its pixelwise multiplication with the wet snow index that yields the integrated snow index. Additionally, some of the analysis and discussion could use some more consideration. Finally, in terms of writing, while I enjoyed the style, I think this paper might benefit from another couple of skims to address issues related to clarity, verb tenses, unnecessary initialisms, misspellings, etc. I’ve provided more detailed comments and grouped them by subject below. For these reasons, I recommend revisions before publication.

Response:

Thank you for the constructive and thoughtful feedback. We greatly appreciate your recognition of our contributions to SAR-based wet snow mapping, as well as your positive comments on our methods and visual presentation. We are committed to addressing the methodological concerns you raised and clarifying the motivations and interpretations of our proposed metrics. Your insightful discussion of the potential limitations of our method has been especially valuable in improving the manuscript. In response, we have added more detailed explanations of our methodology and clearer interpretations of our proposed metrics. Additionally, we have extended the discussion section to explicitly address the limitations of our approach, including the challenges associated with glacier surfaces, the influence of TSI on classification, and the need for dynamic thresholding in larger-scale applications. These revisions aim to enhance the clarity, coherence, and overall quality of the paper. Thank you again for your valuable feedback and for helping us refine our work.

Specific comments:

Methods

Line 128: Do you follow the full Nagler et al., 2016 method, or just the steps listed in your section 3.1? Might want to think about the implications of evaluating against Rc instead of the complete Nagler method

The method proposed by Nagler et al., 2016 primarily involves three key steps: (1) preprocessing VV and VH backscatter images for data preparation and creating a reference image by averaging multiple snow-free images, (2) calculating the backscatter ratio between each image and the reference image for the VV and VH channels and combining these with LIA-dependent weighting factors to derive the composite ratio (Rc), and (3) applying a threshold of -2 dB to Rc for distinguishing wet snow from dry snow.

In our approach, we adhered to these very same steps. As this paper does not aim to reintroduce the full details of Nagler's method, which is well established within the community, we did not elaborate extensively on the Rc generation process. However, Section 3.1 includes the necessary processing steps and parameters to illustrate how we generated the backscatter ratio. We also emphasized our specific selection of reference images in the challenging Karakoram region in the revised text.

For the comparison experiments, we evaluated our method against the threshold-based approach of Rc < -2 dB, ensuring a direct comparison with the Nagler method.

The revised text emphasizing the choice of reference images are as follows: "*Note that this differs from other alpine regions, such as the Alps, where summer months are often used as the reference due to snow-free conditions. In the Karakoram, using winter images as a reference effectively highlights the contrast in backscattering intensity between summer wet snow (lower intensity due to water absorption) and winter dry snow.*"

Line 133: Maybe find some way to motivate the introduction of Rc, or at least give a more thorough explanation of what it is and why we are using it, and why it might be better than VV or VH alone. Perhaps consider adding a line about why a weighting factor W can be helpful in the first place… local incidence angle (LIA) dependence of wet snow detection, etc.

Thank you for the suggestion. The concepts of the composite ratio Rc and the use of weighting factors (W) based on local incidence angle (LIA) are well-established and detailed in the Nagler et al., 2016 paper. In the revision, to enhance the clarity of our paper, we added a brief explanation and motivation for the use of Rc and the benefit of LIA-dependent weighting at the beginning of the section as the following:

"*With the preprocessed γ^0 images, we derived the composite backscatter ratio Rc following the method proposed by Nagler et al., 2016 . The Rc metric combines the backscatter ratios from both the VV and VH channels to comprehensively assess surface condition changes associated with snow melting. This approach incorporates a weighting factor W, which is determined by the local incidence angle (LIA, θ), to account for variations in backscatter due to differing incidence angles. This adjustment enables more robust snow melt detection across varying terrain geometries*".

Line 137-139 / 153-154: The wet snow index is a bit unclear to me. Why is Rc insufficient? What is WSI exactly, and what are you claiming it represents? It is not clear to me why you cap the WSI to 10… if WSI were capped at 1, given two Rc values, I see how the absolute differences between their WSI values would be scaled, but don't the relative distances between their two WSI values remain the same? I guess I don't follow how this allows larger Rc variations.

Thank you for your comments and insightful question. While Rc alone effectively reflects surface condition changes, it can be sensitive to local variations and lacks adaptive boundaries for wet snow classification. Rondeau-Genesse et al. (2016) suggested using a logistic function to convert Rc into a probabilistic space, allowing a soft threshold for wet snow classification. However, this approach requires empirical parameter selection for the logistic function and threshold values. In our work, we utilized GMM to dynamically determine the logistic function parameters, transforming Rc into the Wet Snow Index (WSI) to represent the probability of wet snow presence.

A logistic function typically outputs values between 0 and 1 with the carrying capacity L=1 (indicating probability). Here we chose to rescale the WSI range to 0–10 by setting L = 10 to allow a better combination with the Topographic Snow Index (TSI) used in later steps. This rescale can amplify the distinction between two Rc values. For example, if two points originally had WSI values of 0.1 and 0.9, the rescaled WSI values would be 1 and 9, thereby showing larger discrepancies.

We added the following sentences in the revised manuscript to address the motivation of replacing Rc with the WSI to line 137-139: "*While Rc alone effectively indicates surface condition changes, it can be sensitive to local variations and does not inherently incorporate adaptive boundaries for wet snow classification. Therefore, instead of directly applying a threshold to Rc for wet snow classification, we propose using a GMM to convert Rcinto a Wet Snow Index (WSI) to have a probabilistic measure that better captures the varying conditions of wet snow across different terrains. By leveraging the density distribution of Rc values learned through the GMM, the WSI enables a dynamic scaling of the classification based on the underlying distribution of Rc values.*"

To address the second concern regarding the re-scaling of the WSI values, we re-wrote equation (5) as:

$$WSI = \frac{L}{1 + e^{k(x - x_0)}}$$

and we rewrote line 154 as: "*where k the slope factor, x_0 the logistic curve's midpoint and L the carrying capacity representing the supremum of the function. Here the carry capacity L was set to 10 to amplify the differences between pixels of different Rc values.*"

Line 155-157: I think you should motivate why the GMM is necessary…GMM adaptively determines k based on cluster separation, but what does that equate to in practice?

The GMM was introduced as an unsupervised model to adaptively determine the optimal parameters for the WSI logistic function. It allows the algorithm to choose the proper parameters for the conversion between Rc and WSI, which better accommodates the local variation of signal caused by challenging surface conditions. The slope factor "k" controls the slope of the curve, thereby controlling the sensitivity of the WSI to the difference between wet and dry (or no) snow.

In the revision, we further clarified the motivation and the purpose of parameter K with the revised sentences in line 155-160 as the following: "*In the WSI logistic function, the slope factor k is determined by the separation between the two clusters, providing a flexible and adaptive control over the sensitivity of WSI to the difference between snow conditions. In the case when the two clusters are perfectly distinct from each other (i.e. $|\mu_1 - \mu_2| \gg (\sigma_1 + \sigma_2)$), the WSI would be transformed into a step function and thus effectively act as a single-value hard-threshold at x_0. Therefore, the Nagler's method can be taken as a special case under this assumption with $x_0 = -2dB$. Contrarily, the mixed clusters (i.e. $|\mu_1 - \mu_2| \ll (\sigma_1 + \sigma_2)$) would lead to progressively flattened WSI and soft segmentation boundaries. A similar form of logistic function was proposed by \citet{rondeau-genesse_monitoring_2016}, which was determined with empirical parameters and was used as soft thresholds to replace the hard-threshold of -2 dB on Rc. In our approach, the GMM allows for an adaptive choice of the parameter k based on the distribution density of Rc, thereby enabling flexible and robust applications in large scale*."

Line 160-172: How do you control for the different wet snow backscatter response at different LIAs / polarizations? Rc combines VV and VH as a function of LIA, which gets at whether VV or VH has more wet snow detection capability at a given LIA. Does using Rc as the basis for WSI muddy the water in some way because each pixel is measuring a different proportion of VV and VH data?

Thanks for this question. Using Rc as the basis for WSI does not introduce confusion in terms of proportionality between VV and VH. The LIA-dependent weighting is applied to combine VV and VH because the two polarizations have different sensitivities to snow wetness at varying LIAs. This weighting allows Rc to effectively capture backscatter changes due to snow conditions while minimizing the influence of incidence angle. Rather than reflecting a "proportion" of VV or VH, Rc leverages the most effective part of each signal based on LIA, resulting in a more robust, LIA-independent representation of backscatter change.

Line 160 / 174-176: I am concerned with TSI and its implications on SI / your final wet snow classification. I understand your scaling intent, but I think you may be biasing your wet snow detection across different topographic bins. If my understanding is correct, you calculate WSI at each pixel for an entire scene, then group pixels into the topographic bins, and then take the median WSI in each bin which gives you TSI. Then you get pixelwise SI values by multiplying WSI by TSI. Does this mean you are implicitly assuming something about the relative proportions of wet snow and dry / no snow in each bin? Consider two scenarios: for a lower elevation bin, assume 10% of pixels have wet snow in the summertime SAR image. The majority of the WSIs should be close to 0, and your TSI will likely be low. For those pixels in the bin that are indeed wet, their WSI value now gets multiplied by a very low TSI, and now has a much greater chance of going undetected as wet due to a low SI. Now instead consider the case for a much higher elevation bin, in which the summertime SAR image shows 80% of pixels containing wet snow. The TSI will likely be closer to 10, and so even a pixel with a WSI of 1.5 will end up with an SI of 15. Given how sharp the TSI changes in Figure 4, this could be concerning because it seems like TSI forces pixels towards the classification of the majority of pixels in their topographic bin, potentially overriding each pixel's individual backscatter response. I believe this can be seen in Figure 7… notice how below 5000m, you almost always are underpredicting snow cover relative to S2 snow cover. This consistent low bias / false negatives at lower elevations would be expected with the issue mentioned above. *I would really like to see a plot of F1 score or other classification metric aggregated by your TSI bins (same axes as figure 4)--this could reveal the influence of TSI on your classification performance*. To address this issue with minimal changes to your methodology (if you are truly confident in your method despite these concerns), you might consider some sort of figure that shows spatial maps of Rc, WSI, TSI, SI, SI classification, and S2 snow cover map all included, ideally all with histograms? So something similar to figure 5 but adding in WSI and TSI maps

and histograms. I think this may help readers build intuition for WSI and TSI, and especially how combining these via multiplication into SI ultimately improves your classification.

Thank you for your in-depth analysis of the TSI and its role in our wet snow classification. You raise a valid point regarding the potential bias introduced by multiplying TSI with WSI, particularly the possibility of false negatives in low TSI bins and false positives in high TSI bins.

We acknowledge that, without a supervised approach, some degree of bias is inevitable when applying TSI to different topographic regions. To address this concern, we have taken specific steps to mitigate these biases in the method design:

- **Use of Median for TSI Calculation**: Rather than using the mean WSI in each topographic bin, we calculate the TSI as the median WSI within each bin. This choice reduces the impact of extreme WSI values within each bin, making TSI more robust to variations and less likely to skew toward outliers, thus minimizing the bias you mentioned.
- **Adjusted Threshold for Final Classification**: We selected a threshold of 3.5 × WSI, rather than the midpoint value of 5 from the TSI range, for our final classification. This adjustment helps control the bias by moderating TSI's influence on WSI in both low and high TSI regions, balancing the potential for overestimation and underestimation in different elevation ranges. More about the choice of 3.5 can be found in the next comment.

We have plotted classification metrics across topographic bins (F1 score, precision, and recall) as you suggested in Figure 1~3 . The results show that, while there is a lower recall in lower elevation regions (indicating some false negatives) and lower precision in higher elevation regions (indicating some false positives), the overall F1 score remains consistent across elevation bands. Additionally, the results demonstrate increased precision in the low-middle elevation range and improved recall at high elevations, which suggests that the multiplication of TSI with WSI achieves a balanced trade-off between overestimation and underestimation.

We have added the following paragraph in Section 5.1 to discuss the potential bias introduced by the TSI:

"*Finally, the use of TSI introduces potential bias in regions where topographic conditions deviate significantly from the assumptions underlying its calculation. For instance, low TSI values in ablation zones may lead to underestimation of wet snow, as noted in comparison with S2 results. While the inclusion of TSI improves overall robustness by integrating terrain characteristics, future studies should evaluate strategies to mitigate these biases, especially in regions with complex topography or unique land surface characteristics*."

We will add relevant figures in the supplement. Thank you again for your insightful feedback, which has allowed us to clarify and further validate our methodological choices.

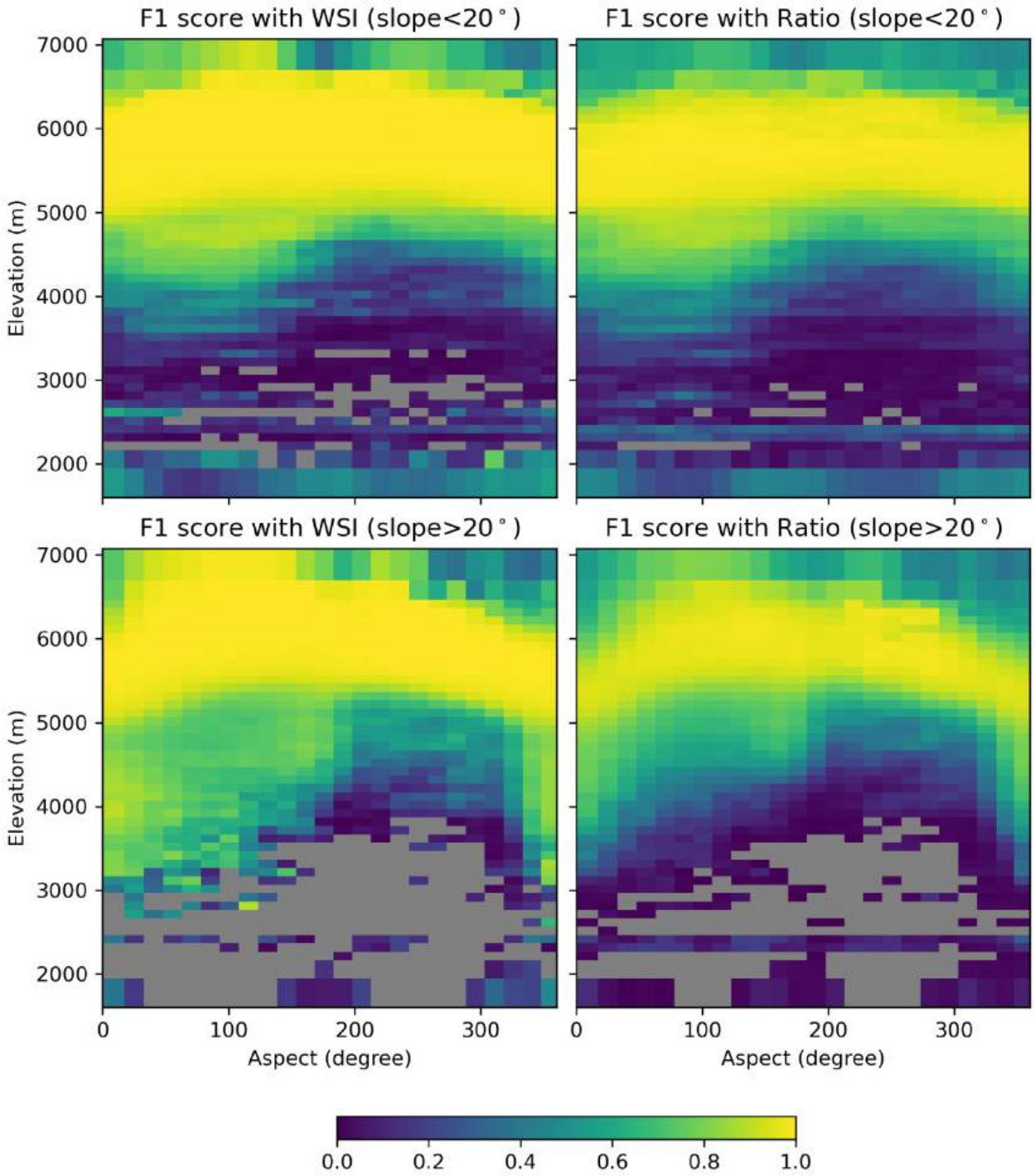


Figure 1. F1 score aggregated by topographic bins in Hunza. Top rows are with slope < 20 degrees, and bottom rows are with slope > 20 degrees. Missing values are indicated with gray.

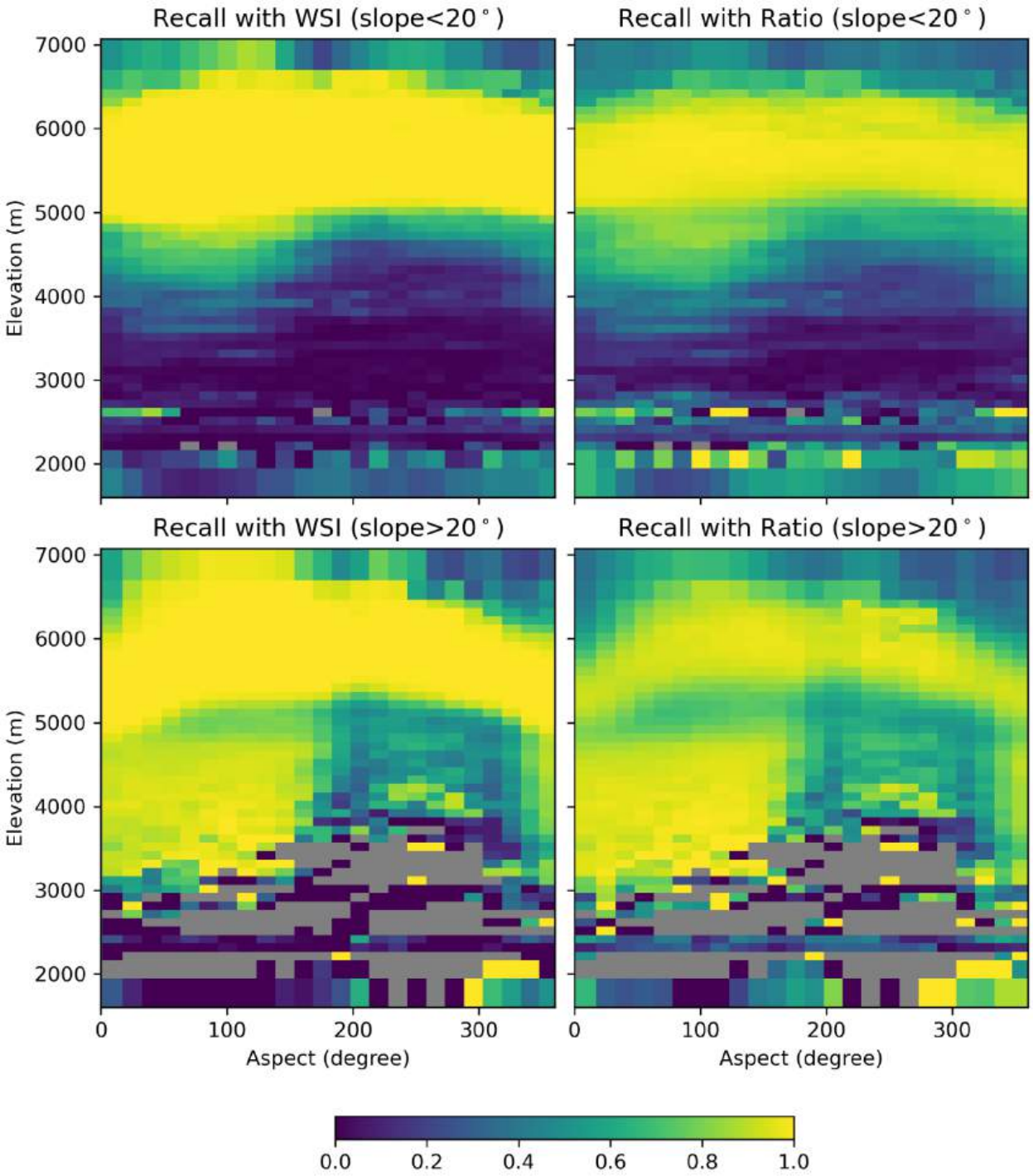


Figure 2. Recall metric aggregated by topographic bins in Hunza. Same arrangement as Figure 1.

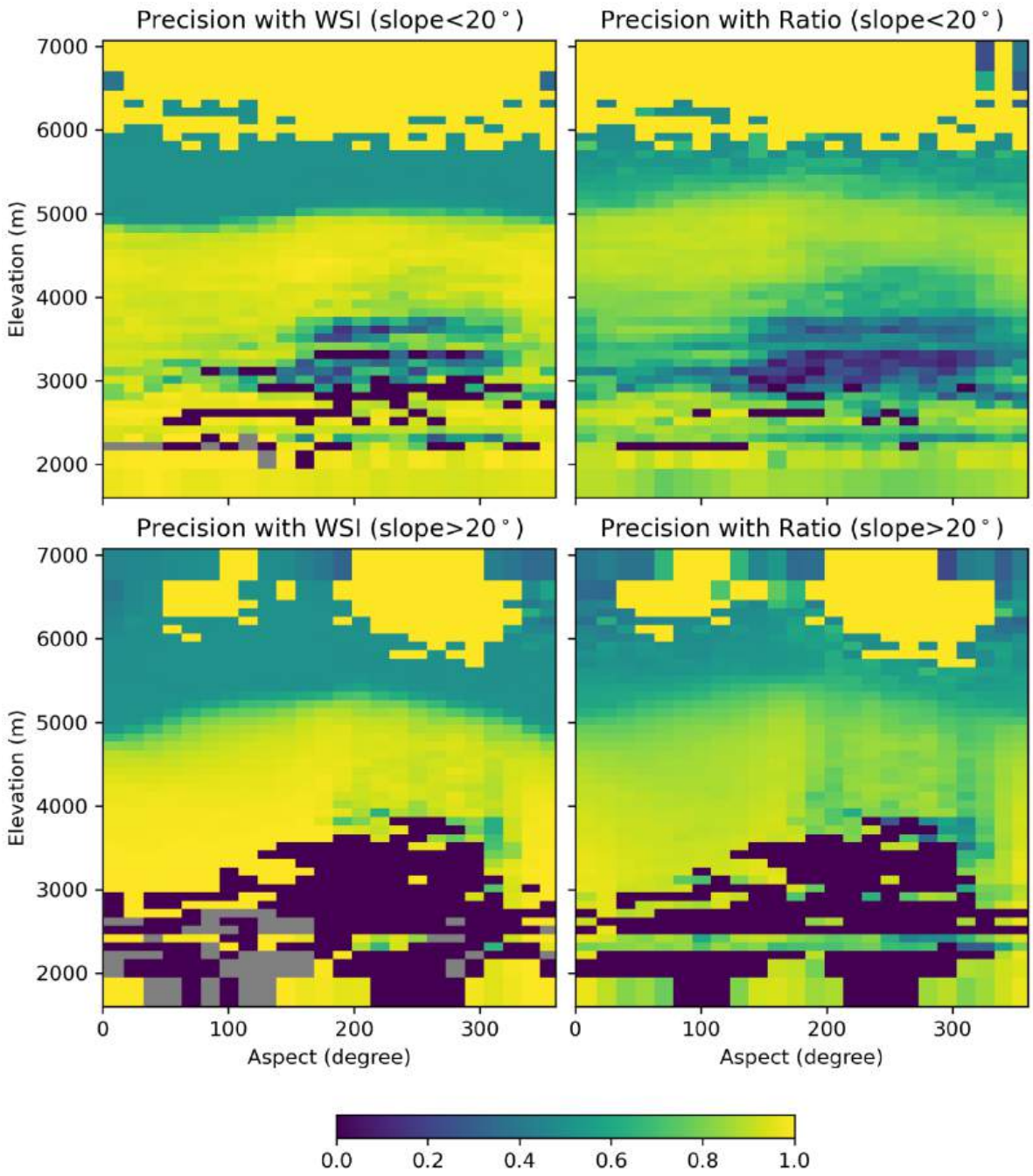


Figure 3. Precision metric aggregated by topographic bins in Hunza. Same arrangement as Figure 1.

Line 179-181: I'm confused on how 3.5 was chosen, it feels relatively arbitrary. Is there an analysis of different coefficients and how they affect the results? Does the coefficient vary across basins, and does it vary annually / based on summer image? If a claim of this paper is to be able to make snow detection algorithms more robust to different locations, will this cutoff have to be chosen dynamically?

Thank you for pointing out the confusion regarding the choice of the coefficient for SI thresholds. To demonstrate the considerations behind the coefficient choice, we will add the sensitivity analysis in the manuscript.

As shown in Figure 4, we evaluated the sensitivity of different classification metrics (F1 score, precision and recall) on the choice of the coefficient. This analysis showed that a value of 3.5 yields optimal metrics across the three basins. Notably, while Hunza and Shyok exhibit similar responses to this coefficient, Shigar reaches its optimum around 2.5. However, since this analysis was based on selected validation dates, we chose 3.5 as an overall coefficient to avoid overfitting.

Regarding seasonal variation, the coefficient does not change annually or based on individual summer images. This is because WSI is derived using the GMM trained on randomly selected samples from all summer images. Thus, WSI represents an aggregated measure for each basin across observed years, making TSI an overall topographic representation rather than a season-specific one.

As for dynamic cutoffs, the need depends on the application scale. In our study, the three basins are geographically close but exhibit unique local topographic features. We performed sensitivity analysis to evaluate how the threshold multiplier affects accuracy. The results in Figure 4 show that a basin-specific optimal coefficient could be used if overfitting were not a concern. For larger, more topographically diverse areas, we anticipate that a more dynamic cutoff would be necessary. One approach could be to apply supervised learning models, such as random forests or neural networks, to capture the non-linear relationships between SAR and topographic features. While this is beyond the scope of the current study, we intend to explore these methods in future research.

To better explain the choice of the coefficient, and discuss the limitation of our method, we revised the section 3.4 in the manuscript as the following:

```

*As illustrated in Figure 2, the final step generated an integrated Snow Index (SI) map by performing pixel-wise multiplication of WSI and TSI. This multiplication scales the WSI by incorporating terrain characteristics, thereby linking the observed SAR backscattering ratio directly with terrain properties.*

*In order to classify the integrated SI into binary snow maps, it is crucial to apply an adaptive threshold that accounts for the variation in topographic features across different basins. The variation in SAR backscatter response within a basin is inherently handled by the GMM when deriving the WSI. In contrast, the TSI is time-varying and basin-specific, requiring an optimal coefficient to condition the SI for classification. To determine this coefficient, we performed a sensitivity analysis, evaluating F1-score, precision, and recall across different values using the S2 validation snow map.*

*The results, shown in Figure 5, demonstrate that Hunza and Shyok exhibit similar responses, with optimal coefficients close to 3.5, while Shigar reaches its optimum at approximately 2.5. However, to avoid overfitting to specific basins or validation dates, we*
```

selected 3.5 as an overall coefficient to balance classification performance across all basins. This coefficient also reflects a moderate threshold applied to the TSI to determine the overall SI threshold for each basin.

The threshold is calculated using the following equation:

$$SI\ threshold = 3.5 \times WSI|_{R_c=-2}$$

where $WSI|_{R_c=-2}$ represented the WSI at a backscatter ratio Rc = -2dB for each basin. This value is basin-specific, allowing the threshold to adapt based on each basin's distinct characteristics. Together, these conditions form an integrated, basin-adaptive thresholding mechanism, combining SAR backscatter and topographic information into a single index to determine the SI threshold.

It is important to note that while the SI threshold is basin-specific, it is time-independent. The WSI is derived from a GMM trained on samples collected from multiple summer scenes over several years, ensuring that it represents an aggregated measure for each basin and is not tied to individual scenes or seasons. This design ensures robustness to seasonal variations in liquid water content and enables consistent application across different validation dates.

```

We also added an discussion about the limitation of the dynamic cut-off in Section 5.1 as follows:

```

While the proposed method demonstrates strong performance across the three basins in Karakoram, several limitations should be acknowledged. First, the choice of the SI threshold relies on the coefficient determined through sensitivity analysis for the selected study areas. While this approach balances classification performance across the basins, it is not fully dynamic and may require adaptation for larger, more topographically diverse regions, such as global-scale applications. Future work could explore supervised learning models, such as random forests or neural networks, to capture more complex, non-linear relationships between SAR backscatter, topography, and snow conditions, enabling dynamic threshold adaptation.

```
```

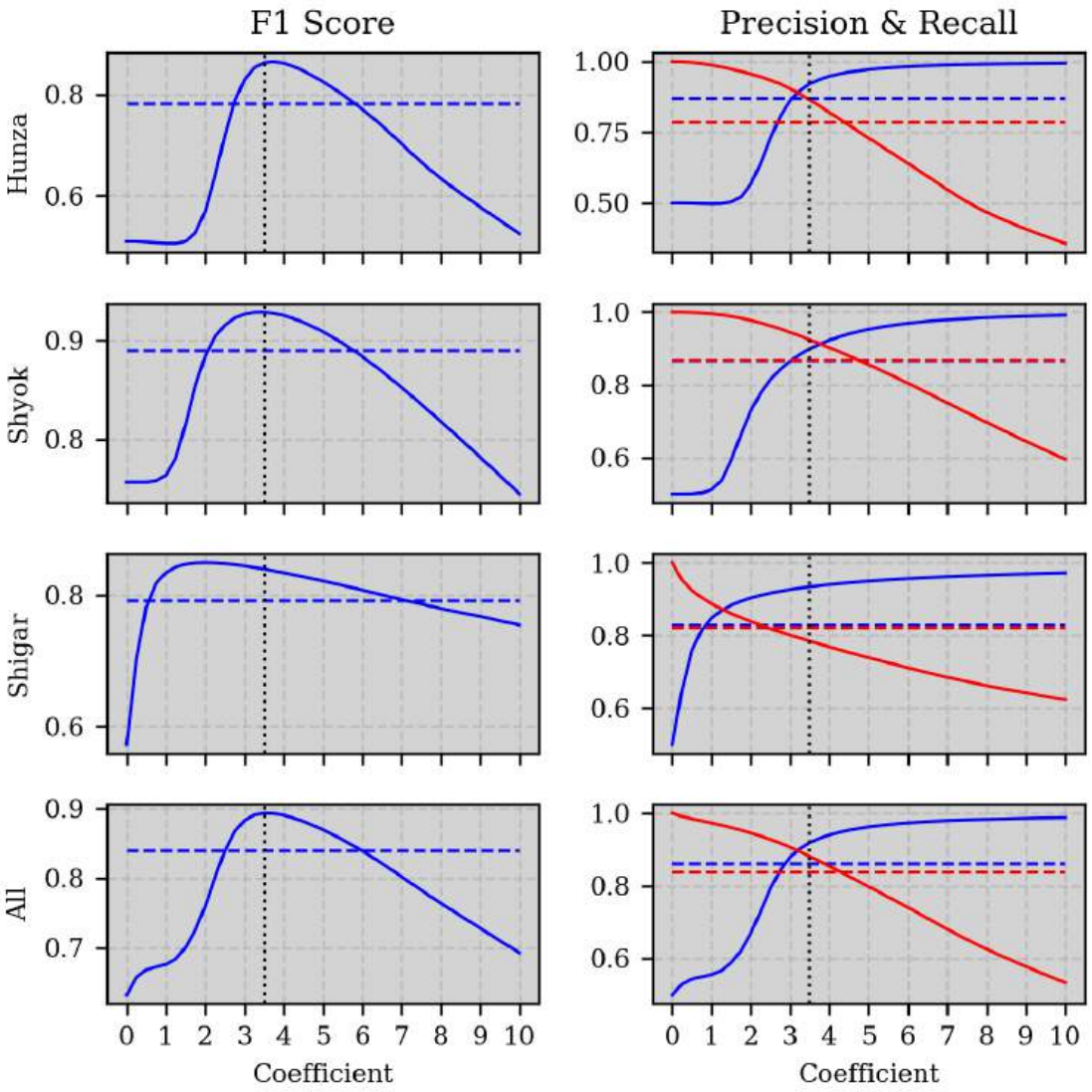


Figure 4. Sensitivity analysis on the choice of coefficients for SI thresholds. The left column is for F1 score, and the right column is for Precision (blue) and Recall (red). Dashed lines are results using Rc=2dB as threshold. The dotted vertical line is 3.5.

Line 268-270: Hmmm, can you just scale to 365 like this? Is this a common practice? I would be concerned about sampling, I'm sure there might be a better way to quantify SMD, maybe day of (last day of wet snow - day of first wet snow)?

Thank you for raising your concern about this point. In the collected dataset, we had a varying number of observed days for each year, resulting in a time-series with uneven intervals. If we use the last day minus the first day, we can not guarantee that the first and last day in the collected data are indeed the exact first and last day. When we scale the observations on a 365-day annual basis, the observation period are standardized to an annual basis and therefore allows for consistent comparison between different years.

We revised the paragraph as follows to elaborate on this point:

"*The SMD reflects the temporal persistence of wet snow cover within a given year, enabling consistent comparisons across years with varying numbers of observation days. To compute the SMD for each year, we first determined the ratio of days with wet snow cover (M) to the total number of observed days (N) for each pixel. Since the number of observed days (N) varied each year and was typically less than 365, we rescaled this ratio to a 365-day basis using the formula (M/N) ×365. This standardization allows us to calculate the annual average of wet snow cover days, facilitating consistent comparisons between years*."

Interpretation of results

Table 2: The adaptive SI thresholds seem to be quite different for these three study sites that are relatively close to each other. It would be helpful to see more interpretation on this.

As noted in our response to the comment on Line 179-181, we will include the sensitivity analysis as shown in Figure 4 to clarify the choice of SI thresholds across the basins.

Line 203-206: This gets at two important limitations… it is possible at the date you select, not all snow pixels are wet, and even across scenes where snow is wet at all pixels, your choice of summer date likely influences how much liquid water is in the snowpack and therefore backscatter response (Baghdadi et al., 2000; Karbou et al., 2021). My guess is that the backscatter drop is more pronounced during certain portions of the melt season. Given this, I think it is important to see how your adaptive SI thresholds change with your choice of summer scene.

Thank you for your thoughtful comment. We agree that variations in liquid water content in the snowpack can impact backscatter response. To make the WSI robust against these changes, we fitted the GMM for each basin using samples taken from multiple summer scenes across several years, thereby determining stable parameters for the WSI. This ensures that the WSI is specific to each basin but independent of the particular summer dates chosen for validation.

Regarding the TSI, while it varies over time, we set a fixed value of 3.5 in the SI threshold calculation, resulting in an SI threshold of $3.5 \times WSI_{Rc=2}$. This threshold is basin-adaptive but time-independent, enabling seasonal robustness within each basin. Consequently, our basin-adaptive SI threshold remains consistent regardless of the chosen validation scenes, and the influence of liquid water variability on backscatter response is inherently addressed through the multi-year GMM fitting.

In the revision, we added the following sentences for clarification: "*In the three basins, adaptive SI thresholds were used to generate the snow classification maps. The threshold*

values for each basin are also reported in Table 2. These thresholds provided basin-adaptive and time-independent classification boundaries to distinguish wet snow and dry snow or snow-free pixels."

Line 211-213: It might be helpful to mention that the noisy patterns over the glaciers are harder to ascribe to the variables you list because of glacial movement. Not sure of the glacier speeds in this area, but it is possible that you are seeing large backscatter changes due to glacier movement that occurred between your winter scenes and summer scene.

Thank you for this insightful suggestion. We agree that glacier movement could contribute to the observed noisy patterns in the Rc map over glacier surfaces. In addition to the mixing of wet snow, water, melting ice, and vegetation within a pixel, glacial movement between the winter and summer scenes may cause additional backscatter variability, further increasing uncertainty in these regions.

In the revision (Section 4.1), we added more interpretation regarding results over glacier surfaces. Please find detailed discussion and the revision in the reply to the comment on Line 217-219.

Line 213-215: I'm not sure it's correct to claim SI reduces the noisy pattern because of the rescale to WSI, I think it's just the multiplication of TSI which is essentially smoothing the pixels in those particular bins, right?

Thank you for this observation. We have revised the sentence to clarify this point as you suggested:

"*On the Integrated SI map, the noisy pattern was greatly reduced. This improvement is primarily attributed to the incorporation of topographic information through TSI, which smooths SI values within each topographic bin.*"

Line 216: You could consider showing these distributions you mention, it would be really helpful to contextualize SI ranges.

Thanks for the suggestion. We added the following histogram of Rc, WSI, TSI and SI in the three basins in the revised manuscript.

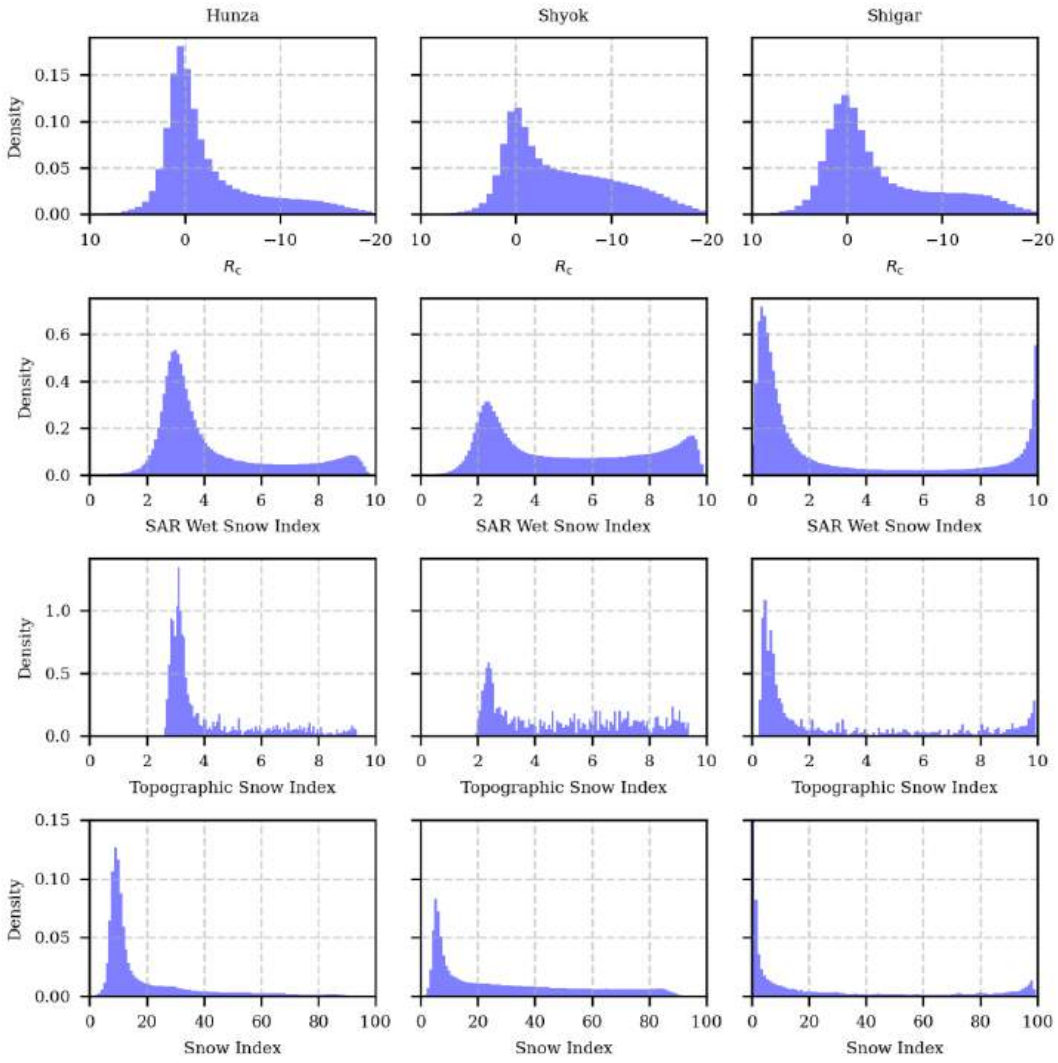


Figure 5.Distribution of Rc, WSI, TSI, and SI of the three basins (left: Hunza, middle: Shyok, right: Shigar). The x-axis of Rc is inverted to align with the indexs (negative Rc corresponds to higher probability of wet snow).

Line 217-219: I'm not too sure you can say this is effective over glacier surfaces… you predict very little wet snow on glaciers, but wouldn't we expect the glacier surface to have a significant amount of wet snow or wet ice in the summer, which should significantly drop the backscatter and be classified as wet snow (even though technically it could be wet ice)? For SAR identification of glacier melt, I would check out Scher et al., 2021. I am concerned about this, especially because in 5c not a lot of glacier area in the insets seem to be identified with wet snow. I think the TSI plays a large role in reducing the SI in these locations (especially with the two slope bins). It may be worthwhile to consider removing glaciers from your analysis, due to these reasons and the fact that significant glacier motion has likely occurred between the winter and summer scenes, so the area in a pixel in your reference image is not the same area in the same pixel in your summer image.

Thank you for your detailed feedback and for highlighting Scher et al., 2021, which provides valuable insights into the complexities of glacier surface scattering. We agree that glacier surfaces represent a challenging scenario for wet snow detection and that further investigation is needed to evaluate how the SI responds to glacier-specific conditions.

Based on Scher et al., 2021, the scattering response in the ablation zone is highly complex and not fully characterized. Processes such as radar backscatter brightening during refreeze and the presence of supraglacial features (e.g., crevasses, suncups, debris cover) contribute to highly variable scattering mechanisms over short distances. These complexities, combined with snow-off conditions, likely result in significant backscatter variability that is challenging to model.

In our comparison between S1 and S2 results, we treated wet snow and wet ice/water as distinct classes. The S2 map incorporates a threshold on the NIR band to separate ice/water from snow. For example, in the insets of Figure 5(d), large areas of glacier ablation zones are classified as ice/water. These regions correspond to snow-free classifications in Figure 5(c). Based on the S2 result, we interpret these areas as being under snow-off conditions, where SI correctly identifies them as snow-free, even though they may contain wet ice or melting glacier surfaces.

We acknowledge that the TSI plays a significant role in reducing SI values in these locations. Ablation zones typically occur at lower elevations with flat slopes (e.g., slope bin 1), which are associated with lower TSI values in the summer. The TSI effectively incorporates topographic controls into the algorithm, reflecting the reduced likelihood of snow persistence in these regions. However, glacier surfaces differ significantly from other land surface types within the same topographic bins. This lack of explicit control for land surface heterogeneity in our method may contribute to the underestimation of snow pixels on glaciers. Future work may explore approaches to better incorporate land surface types into the algorithm, potentially addressing the challenges of glacier-specific scattering mechanisms.

To reflect these points, we have revised the interpretation in lines 217–219 as follows:

"*… The Rc map over these regions present noisy patterns, likely due to the complex scattering mechanisms on glacier surfaces. Over glacier surface, especially in the ablation zone, radar backscatter responses are highly variable due to refreezing, supraglacial features (e.g., crevasses, suncups, debris cover) and the presence of wet debris or bare ice (scher_mapping_2021). These features contribute to significant spatial variability in backscatter within a single pixel, making it challenging to distinguish dry snow, wet snow, and ice with Rc alone. Additionally, glacier movement between winter and summer scenes introduces further variability, compounding the uncertainty in detection*."

We also added a paragraph in the discussion (Section 5.1) to address the challenge and limitations of the proposed method over glacier surfaces:

"*Second, glacier surfaces present unique challenges for SAR-based snow classification. As shown in results, glacier-specific scattering mechanisms, including contributions from wet debris, bare ice, and supraglacial features, introduce variability in radar backscatter that is*

not explicitly modeled in the current method. This limitation may lead to underestimation of wet snow on glacier surfaces and highlights the need for further refinement of the method to better handle glacier-specific conditions, potentially by incorporating land surface type information."

Figure 8: Really pretty plot! For each elevation band, I'm guessing WSE is (wet snow pixels / total number of pixels), not (wet snow pixels / total number of snow pixels), right?

Thank you for your recognition. The WES in the plot was calculated as the total wet snow-covered pixels in a topographic bin divided by the total number of pixels in the respective bin. We will add clarification in the figure caption as the following:

"The WSE was calculated as the ratio of wet snow-covered area to the total area within each respective elevation band."

Line 245-288: It would be great to know how sensitive the WSE and SMD analysis is to the choice of your baseline winter composite and summer image. Do you apply this algorithm to one image and use those SI thresholds? Or do you do it separately for every image in the time series? If it is every image separately, it would be great to know how your WSI/TSI/SI distributions change, and how your SI threshold changes.

Thank you for your thoughtful comment.

The choice of winter composite affects only the Rc calculation. As noted in the introduction (line 43-45), there are several approaches for creating the reference image for ratio calculation. Conducting a full sensitivity analysis on each possible method would shift the focus of this paper from the proposed method to evaluating the optimal reference image composition. Instead, we followed established practices in the literature and selected the winter average as the reference, which is widely accepted for producing a robust Rc calculation.

Regarding the choice of summer images to fit the GMM, we used samples randomly selected from all summer images (July-early September) in multiple years for each basin. This allows the GMM to learn the general pattern of Rc distribution within a basin across the years. Consequently, the WSI and the derived SI threshold, calculated as 3.5×WSI_Rc, are both basin-specific and time-independent, making it robust for use across the full time series.

The TSI, on the other hand, is time-dependent, as it shows the topographic control of snow accumulation and melting in different seasons. In the revision, to illustrate the seasonal variation, we provided example TSI distributions for different seasons in Figure 4 in the manuscript. However, visualizing TSI distributions for every date in a time series would be challenging and, given the stability of the basin-specific threshold, is unlikely to add significant insights.

Clarity

Figure 1: Are the black boxes scene footprints? Maybe make this explicit, and if you keep the boxes, maybe label them or color code them with relative orbit information and add a legend? And I think your colorbar may be flipped?

Thank you for pointing out the problems. In the revision, we have updated the figure with the correct color bar and add color code to the boxes. The boxes are described in the caption as the following:

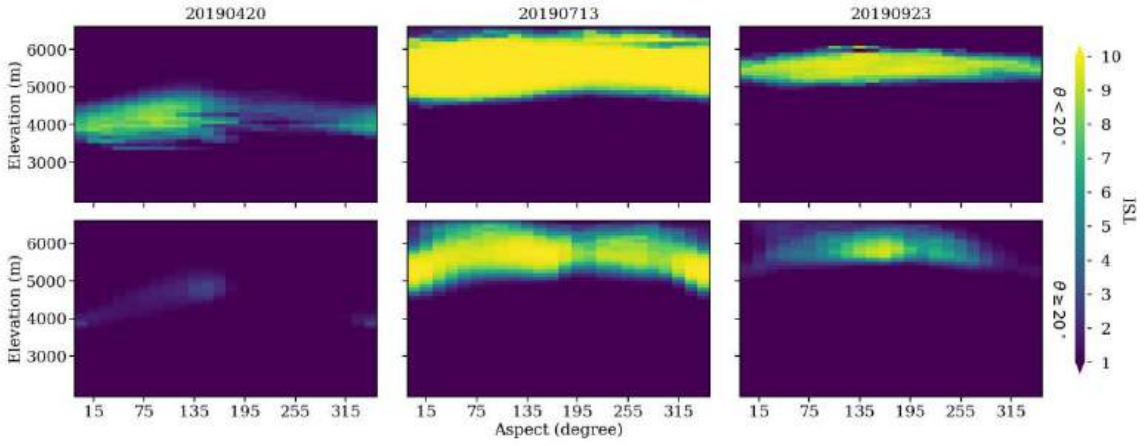
"*Footprint of S1 images used in the study are show with black and blue boxes (black: relative orbit 27; blue: relative orbit 129).*"

Line 81: I'm not too sure what this sentence is getting at, perhaps reword?

This sentence has been reworded as: "*The Karakoram is situated upstream of both the Upper Indus Basin and the Tarim River Basin*."

Figure 4: Not sure if a red and blue diverging color map is most appropriate here. I feel like the white values don't play an especially important role (a value of ~5 for TSI doesn't seem particularly important), and the color map could cause confusion with the reader thinking it is the same quantity as Figure 5's Rc plot (though I think the red and blue color map for that Rc plot is great!).

Thank you for the suggestion. We have changed the color bar to Viridis as shown below.



Line 179: Clarify "…where 3.5 implied the condition applied to the TSI…"

We will revise this sentence as the following:

"In this equation, the factor 3.5 was selected based on validation set results, optimizing for precision, recall, and F1 score. It reflects a moderate threshold applied to the TSI to determine the overall SI threshold for each basin. The term $WSI_{Rc=2}$ represented the WSI at a backscattering ratio Rc at -2~dB for each basin. This value is basin-specific, allowing the threshold to adapt based on each basin's unique characteristics. Together, these conditions form an integrated, basin-adaptive thresholding mechanism, combining SAR backscatter and topographic information into a single index to determine the SI threshold."

Line 225: What percentage of your pixels were excluded, and why was 5500 m a.s.l. chosen?

The excluded percentages were as follows: Hunza 11.08%, Shigar 12.95%, Shyok 30.43%, with a total exclusion of 19.60% across all basins.

We selected 5500 m a.s.l. as the upper limit because areas above this elevation are consistently considered non-melting zones due to low air temperatures throughout the year, which inhibit significant snowmelt. This threshold aligns with the general knowledge about the temperature-elevation relationship in the Karakoram, where elevations above 5500 m are predominantly characterized by accumulation zones or permanent snow cover. By excluding these areas, we focus the analysis on zones where snowmelt dynamics are more relevant to our study.

To clarify this point, we have revised line 225 as the following line in the manuscript: "*Areas above 5500 m a.s.l., representing 11.08% of pixels in Hunza, 12.95% in Shigar, and 30.43% in Shyok (19.60% in total across all basins), were excluded from the calculation. This elevation threshold was chosen to minimize errors caused by the presence of dry snow at high altitudes, where consistently low air temperatures throughout the year inhibit significant snowmelt.*"

Line 237: I would clarify what you mean by "…comparative sensitivity of the SAR signal to dry snow in both methods."

Thank you for your comment. Here, "comparative sensitivity of the SAR signal to dry snow" refers to the fact that both Rc and SI methods show limited sensitivity to dry snowpacks, resulting in decreased snow coverage in the snow maps for areas predominantly covered by dry snow. This reflects a limitation of SAR-based wet snow mapping methods when applied to regions with extensive dry snow, as the backscatter signal from dry snow is generally difficult to distinguish from other surface conditions.

To clarify this point, we have revised the sentence as follows: "*Above 5500m a.s.l, where expansive dry snow cover predominates, both Rc and SI maps showed a greater reduction in wet snow coverage compared to S2, highlighting the differing sensitivities of SAR signals to dry snow conditions in these methods compared to S2's multispectral data*."

Figure 5: It is not clear where the S2 snow cover map is coming from. Is this output from the let-it-snow algorithm? If so, where is this ice/water class coming from? Also, it might be helpful to say in the caption that these are all three basins.

Thank you for your comment. The S2 Snow Cover Map in Figure 5 was generated using the Let-it-snow (LIS) algorithm, as described in Section 3.5. We also updated this section to clarify the LIT algorithm and the generation of ice/water class.

We will revise the caption of Figure 5 as follows:

"Figure 5. (a) Rc, (b) Integrated SI, (c) SI Classification Map, and (d) S2 Snow Cover Map (as reference) for all three basins (Hunza, Shigar, and Shyok). The S2 snow map was generated using the Let-it-snow (LIS) algorithm, as described in Section 3.5. Zoomed insets provide a closer view of selected locations in each basin, highlighting performance differences on glacier surfaces."

Line 289-291: Definitely a really neat approach, but maybe temper your claim of great improvement here, or provide the specific stats.

We have changed “greatly” to “effectively”.

Line 301-304: I think this is a more fair interpretation of TSI! It should be helpful for understanding aggregate backscatter response of topographic bins, I’m not sure about the added use of using it as a scale factor for WSI.

Thank you for your positive feedback on our interpretation of TSI. We agree that it is valuable for understanding the aggregate backscatter response within topographic bins, providing insights into how topography influences snow presence. Using TSI as a scale factor for WSI allows us to incorporate this topographic influence directly into the SI calculation. This approach ensures that regional topographic variability, which can affect snowmelt dynamics, is accounted for in the final classification.

Line 316-318: Provide citation

We added Cluzet et al., 2024 in the revision as reference work.

Line 324: You haven’t previously talked about mitigating uncertainty, so I would avoid introducing this idea in the conclusion.

We avoided using the word “mitigated uncertainty” and reformulated the sentence as the following: “*We first adopted the GMM to adaptively transform the SAR backscattering ratio Rc into the WSI as a robust representation of wet snow under complex surface conditions*.”

Line 324-325: I feel like the purpose/definition of TSI shifts around a bit in this paper. Is it supposed to take into account the influence of topography on snow presence or the influence of topography on the backscatter response of wet snow?

Thank you for raising this point. The purpose of the TSI is indeed to account for the influence of topography on the presence of snow rather than on the backscatter response of wet snow. While TSI is calculated using the median of the WSI, its primary role is to represent the likelihood of snow melting under specific topographic conditions, not to evaluate backscatter response variations due to topography. We clarified this in the manuscript to avoid any confusion as the following:

"*Then, we introduced the TSI to capture the likelihood of snow presence influenced by topographic conditions*."

Technical corrections

Line 5: For the abstract, I would use "a" instead of "the" in front of both Wet Snow Index and Topographic Snow Index so it's clear to that these are metrics that you are introducing
Line 10-11: Remove "the" in between under and climate
Figure 1: "tree" -> "three"
Line 47: "inlucdes", also check tenses / plural of this line
Line 53: Remove "of"
Line 67: "meting" -> "melting"
Line 67: "watersheds" -> "watershed"
Line 69-72: Make verb tenses agree
Line 89: "snow melting changes" -> "snowmelt"
Line 108: "the" -> "a"
Line 116-118: unclear
Line 122: "backscattering intensity images" -> "backscatter images"
Line 123: "firstly" -> "first"
Line 131: add "is" before "the reference image"
Line 216-217: Check tenses
Line 218: "well" -> "good"
Line 221: Consider rewording "greatly false positives"
Table 3: "corresponded" -> "correspond"
Line 278: Consider "This trend" -> "This change"
Line 280: Maybe avoid "significant", might be confused for an actual test of significance
Line 284: Remove "significant"
Line 305: "was" -> "were"
Line 306: Should there be a dash between the two aspects here?

Line 311: “Management” -> “management”
Line 312: Remove “been”
Line 319: Remove “significant”
Line 320: “observation” -> “observation record”
Line 325-326: Wording of this sentence is a bit confusing, consider rewriting and remove “significant” unless some sort of hypothesis testing was done.
Line 328-329: Remove “such as” since these were the only two variables analyzed in the paper.
Many lines: Check usage of initialisms (GMM, BOA, RMSE, MSE, WSI, SVM, RF, ML, TSI, etc). Often initialisms are introduced multiple times, and some initialisms are introduced and never used (or sometimes used once)
Many lines: Tenses are often inconsistently used; I would skim over to make sure they agree!

We highly appreciate the detailed language and clarity suggestions. We have carefully revised the manuscript to address these points. Specifically, we have adjusted the use of articles (e.g., “a” instead of “the”), made corrections to spelling errors and ensured consistent tense usage throughout the text, removed redundant words and clarified ambiguous phrases to enhance readability, refined the use of initialisms, reworded or removed terms such as “significant” to avoid potential confusion, and ensured uniformity in verb tenses and made other minor edits to improve overall coherence. The revisions can be tracked in the revised manuscript with difference highlighted. Thank you for your valuable feedback, which has profoundly contributed to strengthening the quality of our paper.

References

Baghdadi, N., Gauthier, Y., Bernier, M., & Fortin, J.-P. (2000). Potential and limitations of RADARSAT SAR data for wet snow monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(1), 316–320. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/36.823925

Karbou, F., Veyssière, G., Coleou, C., Dufour, A., Gouttevin, I., Durand, P., Gascoin, S., & Grizonnet, M. (2021). Monitoring Wet Snow Over an Alpine Region Using Sentinel-1 Observations. *Remote Sensing*, *13*(3), Article 3. https://doi.org/10.3390/rs13030381

Nagler, T., Rott, H., Ripper, E., Bippus, G., & Hetzenecker, M. (2016). Advancements for Snowmelt Monitoring by Means of Sentinel-1 SAR. *Remote Sensing*, *8*(4), Article 4. https://doi.org/10.3390/rs8040348

Scher, C., Steiner, N. C., & McDonald, K. C. (2021). Mapping seasonal glacier melt across the Hindu Kush Himalaya with time series synthetic aperture radar (SAR). *The Cryosphere*, *15*(9), 4465–4482. https://doi.org/10.5194/tc-15-4465-2021

Supplement

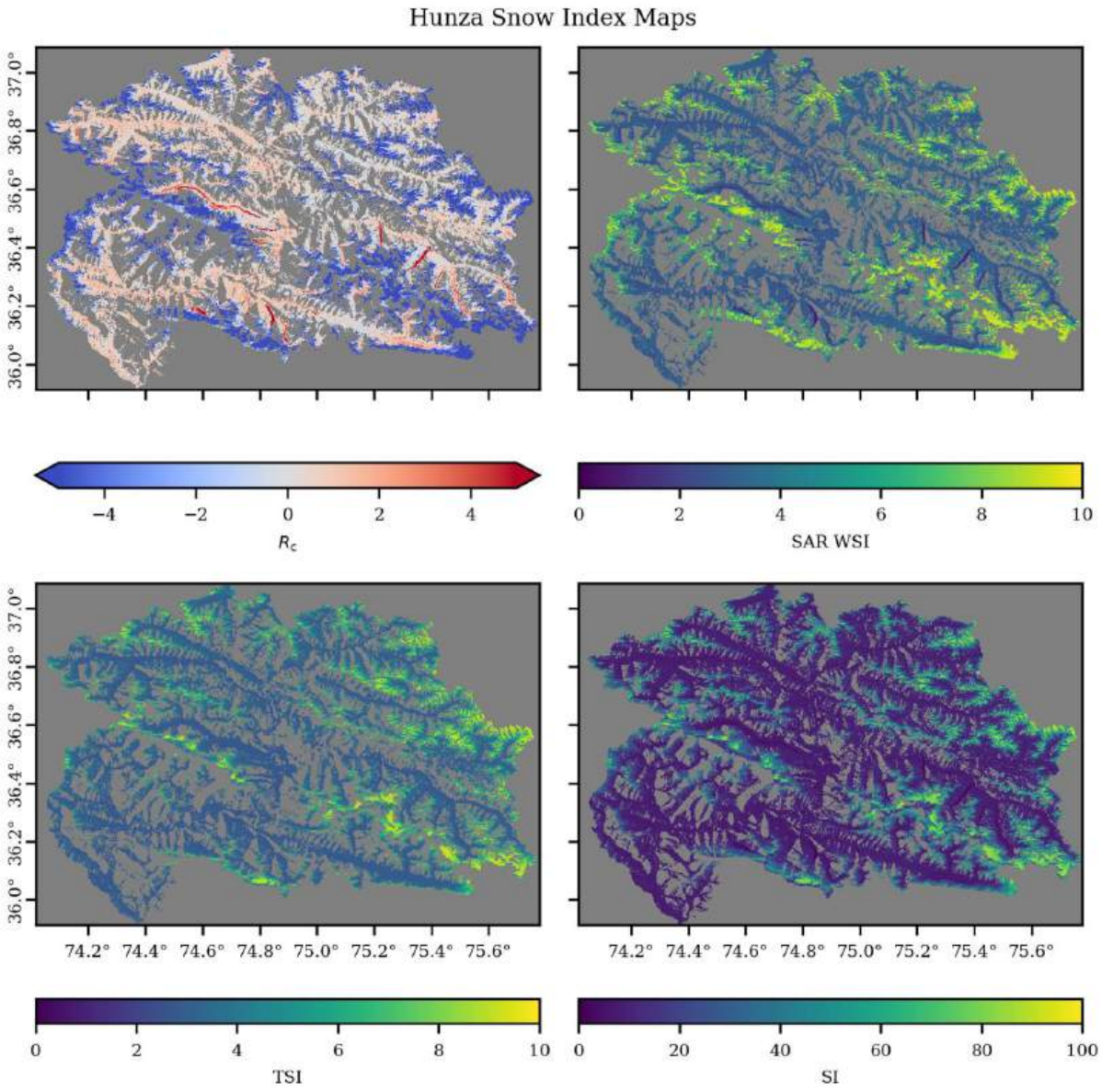


Figure 6. Rc, WSI, TSI and SI maps of Hunza

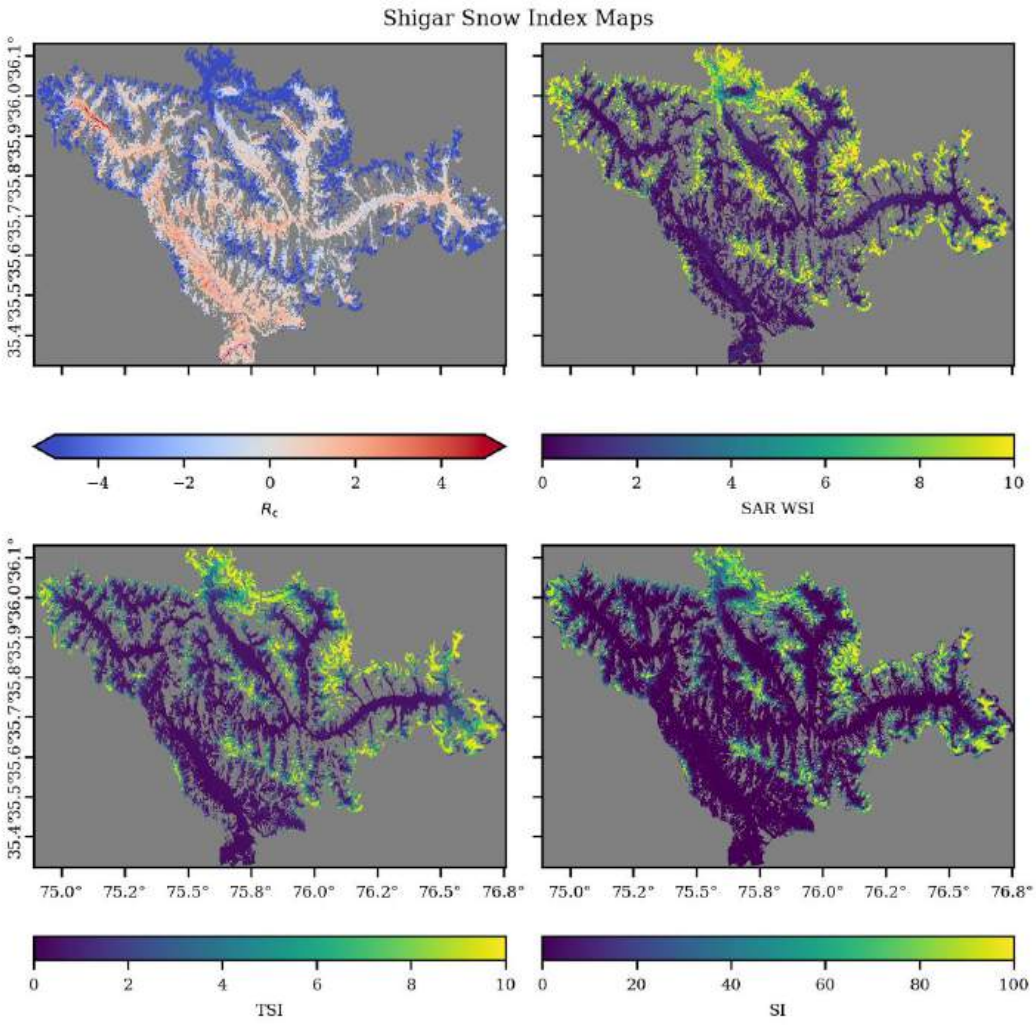


Figure 7. Rc, WSI, TSI and SI maps of Shigar

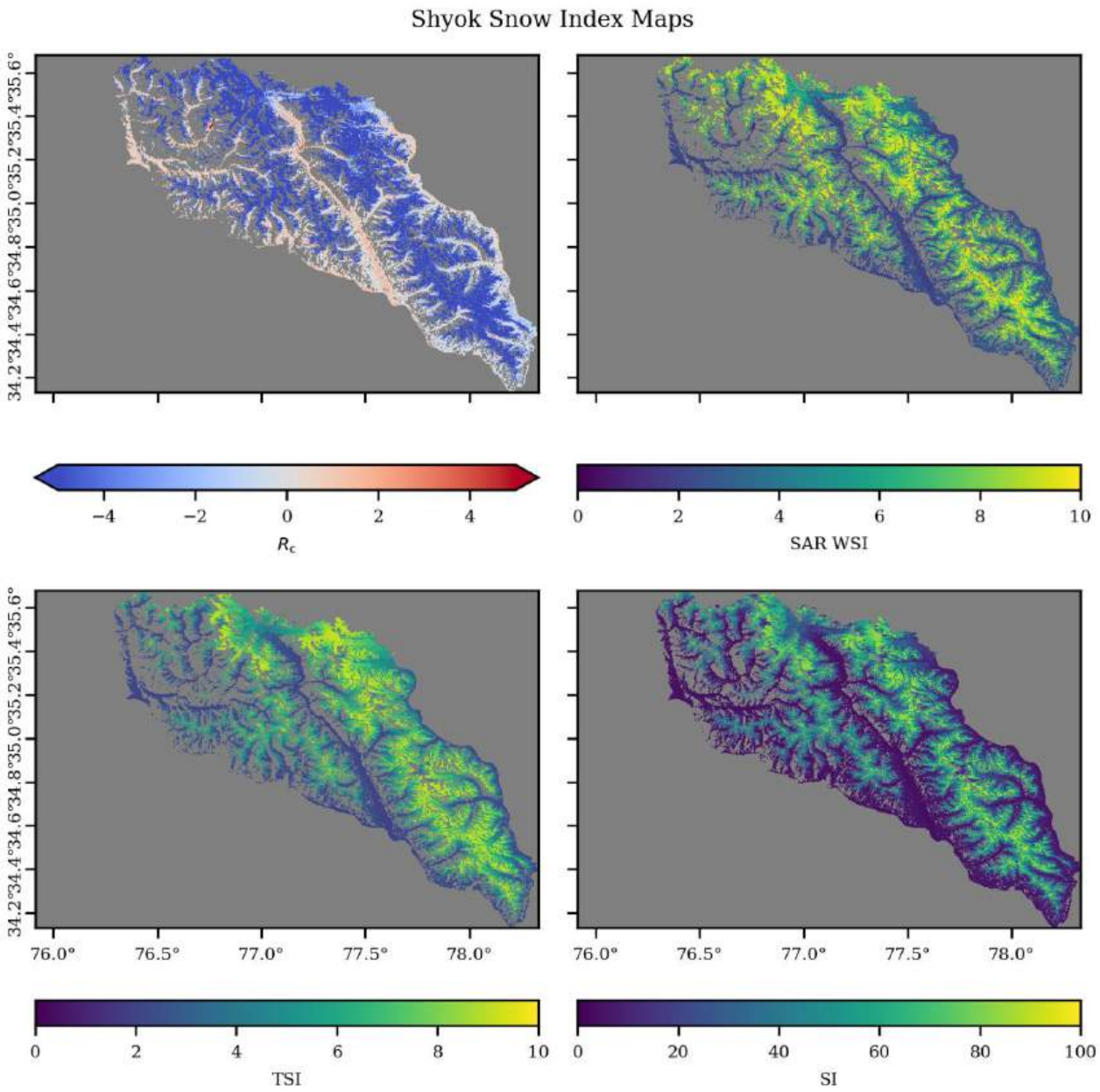


Figure 8. Rc, WSI, TSI and SI maps of Shyok