Review of "Mapping seasonal glacier melt across the Hindu Kush Himalaya with time series SAR" by Scher et al.

This study uses Sentinel-1 SAR data to map the seasonal melt characteristics (melt onset, freeze onset, and melt duration) across the Hindu Kush Himalayas. Using this dataset, the study investigates spatio-temporal variations across subregions and as a function of elevation. Two sets of automatic weather stations are used to validate the interpretation of these studies and highlight where glacier ablation models would perform poorly. Given the lack of in-situ observations and seasonal remotely sensed observations in this region, this type of dataset would be very valuable to modeling glacier melt. The use of systematic C-Band SAR observations in this region also appears to be the first time this has been done now that the satellite imagery is available.

While I believe this would be a valuable contribution to the field, I unfortunately believe there to be several major issues with the interpretation of the dataset. The primary issue being the interpretations of the ablation area, which the authors note is challenging and/or were excluded; although this exclusion is unclear because analyses suggest that all area were included. Furthermore, for a study of this scale, I expected there to be some type of validation to support their findings. Instead, the two sets of automatic weather stations were partially used for validation and partially used to highlight where melt models would perform poorly. I read the author's response to reviewer's where they state that this is not possible, except for the two automatic weather stations; however, I don't agree with this assessment. For example, other relevant studies (some pointed out in the review below), in-situ measurements, and/or other sources of satellite imagery (other microwave datasets, optical datasets, etc.) would have been useful to support the interpretations and conclusions.

In my opinion, there needs to be considerably more validation. I would also suggest that unless the methods can be improved to handle the challenges in the ablation area (which would need to be shown through a rigorous validation process to provide confidence), that the ablation area be excluded entirely, and the results be limited solely to the timing of melt in the accumulation zones. Another alternative is that if debris-covered areas are the only problem, then limit the study to clean-ice glaciers only. Either way would still provide useful information to modelers, although this would greatly reduce the novelty and impact of the study overall. At that point, the novelty of the study would need to be reconsidered. I therefore recommend the paper to be reconsidered after major revisions. Please see my major comments and minor comments below.

Major Comments

The methods show results (e.g., Figure 4 middle; Figure 5). If these datasets were used to develop the method or solely for validation, then this should be explicitly stated. Either way, the assessments of the methods should be in a separate results section. This performance assessment could then focus on validating the methods before the large-scale assessment of spatial trends.

I found the references to support various statements and relevant work needs improvement. I have highlighted many examples in the introduction, but also believe this is necessary for validating the interpretation of the SAR signals as well.

The method does not appear to work over ablation areas. It is unclear from the results presented if this is an isolated issue with debris-covered areas or if it is an issue with clean ice as well because the example figures were only for debris-covered glaciers. Given debris-covered areas are highly prevalent in HKH, this is a major issue. This is clearly apparent from Figure 7 Top. The Khumbu Glacier is debris-covered below the Khumbu icefall. The interpretation of the SAR signal on this debris-covered area is that it indicates refreeze onset between DOY 250-270, which is in September. This is highly inaccurate and highlights the lack of validation within the study. For example, this part of the debris-covered ice is clearly still melting in September (Rowan et al. 2020; *Journal of Glaciology*). Similar issues appear to persist with the Freeze Onset in Figure 8C,D.

It's unclear if similar issues exist for the melt onset signals as well. The color bar in Figure 8A,B is too hard to read to discern if melt onset appears to be happening as early as March 1st, which would be unlikely. However, the fact that the color bar is shows DOY 60 suggests there are some areas where it is melting at this time; otherwise, why stretch the color bar outside the values shown in the figure? The issues with the freeze onset and melt onset and lack of any validation, in my opinion, undermine the entire study.

One recommendation is to put the results and interpretation into better context of glacier zones. For example, the discussion/interpretation primarily focused on the liquid water in frozen snow and percolation faces, but the results that were being interpreted were aggregated from 3000-7500 m.a.s.l. The lower elevations are clearly in the ablation zone, i.e., clean ice or debris-covered ice. Therefore, it's unclear how much of the discussion of the snowpack and percolation faces is warranted. Or is this partially being included to discuss seasonal snowpack covering the ablation zone? Including a rough estimate of the glacier zones, even if they are roughly estimated based on median elevation or some other metric, and their respective hypsometries may provide some useful context for interpreting the results. Otherwise, results/discussion like L406-414 come across as interpreting the entire glacier as being in the percolation zone.

Given the major issues with the ablation area, I would suggest either (i) limiting the study to only accumulation areas, or (ii) limiting the study to only clean-ice glaciers. Either way, there needs to be significantly more validation performed to provide confidence in the results. This validation should ideally be done for both the ablation and accumulation areas, albeit that the accumulation zones interpretation should be much stronger as they are based in theory as the authors clearly discuss in the main text and mention in response to a previous reviewer.

This validation is important as it is unclear how sensitive the results are to various aspects of the methods. For example, is there any sensitivity to the chosen dB threshold of 3? A previous reviewer asked for a sensitivity analysis, but the author's simply responded that this was conservative based on literature values. Even if this is only done for a handful of glaciers, a sensitivity analysis would provide more confidence. Similarly, what about using ascending vs. descending orbits? The choice of orbit direction clearly affects the interpretation of the signal at high elevations (Figure 5) and yet for the full region a composite was used. If this composite is used moving forward, it'd be good to see some sensitivity/error analyses because it appears to have a strong impact on the results, but perhaps this is only for high altitudes.

Similarly, were the high-altitude automatic weather stations used for validation? The energy balance modeling likely has its own issues (e.g., sublimation could be important here?), but the modeling appears to be used to validate the SAR signal (L320). If it is used for validation, does that mean that the SAR signal overestimates the amount of melt by 33-43% (L345)? This would appear to be a considerable amount of error.

Specific Comments

Given there's no word limit, I would suggest writing out acronyms like melt onset and freeze onset to make the study more readable to the average reader.

L34: "Ice caps" have a particular meaning. In HKH, they are primarily (if not all) mountain glaciers or valley glaciers.

L29-46: This introduction bounces back and forth between discussing global issues (e.g., contributing 25% of sea level rise) to HKH specific issues (e.g., freshwater for 2 billion people). It's useful to show how HKH changes fit into the larger picture, but I'd be conscious of this change in scales to make it easier to read.

L47-49: Litt et al. (2019) focuses on two glaciers over a period of a several years. It does not address "projected disturbances". More appropriate reference would be one of the GlacierMIP studies (e.g., Hock et al. 2019 (*Journal of Glaciology*) or Marzeion et al. 2020 (*Earth's Future*)). The same is true for L55, which discusses projections but does not mention relevant studies where these uncertainties exist.

L65-68: Statement concerns inability of studies to capture variability in patterns and magnitude of melting, but only references a study that measures the geodetic mass change (Brun et al. 2017). The more recent study by Shean et al. (2020; *Frontiers in Earth Science*) would be appropriate to reference here as it provides refined estimates of mass change compared to Brun et al. (2017). Furthermore, the recent study on projections in this region (Rounce et al. 2020; *Frontiers in Earth Science*) explicitly captures the variability reported by these measurements using a degree-day model.

L73: "operational monitor" – consider changing "operational monitoring system" perhaps?

L127 – this would be useful for assessing all models of glacier ablation, not just energy balance models.

L137 – for *The Cryosphere* I would suggest using more standard glacier mass change terms (e.g., mass loss, mass balance, mass change) as opposed to the term "wasting" that is less frequently used.

L139-141 – appears to imply that Karakoram, Kunlun Shan, etc. are not affected by increases in global average temperature. I'd suggest references Karakoram anomaly and recent studies finding that they are starting to lose mass (e.g., Farinotti et al. 2020 (*Nature Geoscience*)).

L148 – is there a reference for this ICIMOD dataset?

L158-171 – this is a lot of detail on the glacier outlines used in a study that I don't see being relevant. If debris-covered areas are an issue, there's no mention of datasets that explicitly delineate it (Scherler et al. 2018 (*Geophyiscal Research Letters*); Herreid and Pellicciotti 2020 (*Nature Geoscience*)). This should be discussed.

L189 – "this" should be "the" or "the timeframe of this study"

L204 – suggest stating the name of the cloud-computing platform here, i.e., Google Earth Engine, to make this explicit for readers as opposed to only mentioning it at the end of the paragraph.

L216 – It is my understanding that these are both on Khumbu Glacier. Therefore, this should be "over a high elevation glacier", not glaciers.

L245 – Why is this weather station not stated in Section 2.4?

L247 – Introduction stated that assuming 0 degC threshold is poor assumption and was part of the motivation for this study. May want to preface that this generally works well, except for at extreme altitudes where other processes (e.g., sublimation) may be important (if that is the point that is trying to be stated)?

L296 – It's unclear how debris-covered areas were identified (see previous comment). Nonetheless, not including debris meant between 30-48% of the ice in the ablation area is not being monitored (Kraaijenbrink et al. 2017; *Nature*). It would be good to state how much over the actual glacier area was able to be monitored, and if statements are being made about the ablation area (like that being mentioned here), then stating the percentage of the area in the ablation area being monitored is relevant as well.

L307 – if the debris-covered area is excluded, then how are statements being made concerning those areas? Is the assumption at the regional scale that when evaluating variability, the debris-free areas that were measured at lower elevations are representative of the debris-covered areas? If so, this should be stated explicitly, as it's confusing within its present form. Note that the Khumbu Glacier example clearly shows debris-covered areas being included.

L322 – “are” should be “is”

L324 – If the energy balance modeling is an important part of this study, which it appears to be, then additional detail should be provided in the main text. At a minimum, this should include the values used to force the model (i.e., those in Table SI) as well as relevant information pertaining to the timestep and the terms considered (i.e., Equation 1 in the supplement).

L330-350 – these are results, not methods.

L337-339 – I was thinking this same point, so I’m glad the authors brought this up. However, the timing of satellite overpasses is known (Figure 5 caption), so why was this not performed? This appears to be the primary validation of how well SAR performs, and one of the primary conclusions of the study that SAR can detect melting where models otherwise wouldn’t, so it should be presented in a rigorous analysis to provide confidence.

L351-355 – High degree of confidence from two weather stations, where the days of melting appears to be overestimated by 33-43% (L345) seems to oversell the results. I agree that this is a challenging topic, especially since there are uncertainties with the SAR data and uncertainties with the energy balance model (unfortunately, there’s no validation data for the energy balance model, so are these differences due to the SAR data or the energy balance model – likely both?).

L355 – “HMA” (High Mountain Asia) is not defined in text. It’s used later in text as well. I’d suggest using HKH throughout.

L386 – except for debris-covered areas, no (L297)?

Table 3 – caption and table are opposite directions making it very difficult to read. Also, 1 km elevation bins are very coarse considering that some regions (if not most regions) have most of their glacier area spanning 1-2 km (Figure 4 Bottom). This means that the statistics shown for the other elevation bins is for a tiny fraction of the glacier area, which in my opinion detracts from the overall value of this table. The 100 m bins in Figure 4 are much more meaningful.

Figure 6 – this figure is difficult to read/interpret due to the lines. I would suggest changing both the circles and the squares to lines. The circles and squares are clearly separated, so you could simply add a note that the melt onset is on the left and freeze onset on the right, which is intuitive anyways. This would enable the reader to follow the lines and determine how the trends vary. It will also make it easier to see the 12 subregions, which currently are on top of one another and hard to discern any information from. This will also make it easier to see statements like L400 where it does not follow a linear trend, when the current form of Figure 6 looks like roughly speaking higher elevations are around DOY 250-270 and lower elevations are around 280-330, so while the linear trend is not as strong, there still appears to be a trend.

L393-396 – is this supported by the energy balance modeling?

L429-433 – Khumbu Glacier is highly debris-covered below the Khumbu icefall, which the authors appear to interpret as a signal of refreeze onset. This is highly inaccurate. Figure 7 (Top) suggests the Khumbu Glacier would be refreezing on DOY 250-270 (sometime in September) when the debris-covered ice is clearly still melting (Rowan et al. 2020; *Journal of Glaciology*). Figure 8C,D show similar issues. See major comment.

L533 – Shean et al. (2019a) does not exist. Unclear therefore where the trends for 2000-2010 are coming from and why trends in later decades covering the time period observed (i.e., 2000-2018 from Shean et al. 2020) are not used.

L557 – surface energy balance models "are constrained" by SAR data indicates that the SAR data is used to calibrate the energy balance models. This is in direct contrast to earlier where the energy balance models were being used as validation of the SAR observations (L320).