Anonymous Referee#1

I am pleased to see that the authors now present an improved version of the manuscript including not only a take on a clearer description of the methodology that is now easier to follow, but also providing additional analysis. Nevertheless, I would still recommend that parts of the manuscript be reworded to further improve clarity of method and readability. Below you can find some minor and technical comments.

We thank the Reviewer for the comments that have contributed to improving the clarity of the manuscript. Please find in blue our point-to-point answers.

Specific comments

L5-6: it is not clear to me if you used precipitation data, or SWE/ SD recordings to determine the “accumulation” state (see comments to section 2.1.1). In section 2.2.2 it is stated that the “ablation” state also depends on the degree day (DD > 0).

Thank you for the comment. With the aim to be as general as possible, we meant with “precipitation data” both the data recorded by pluviometers as well as snow depth/SWE sensors. Since the differentiation from liquid to solid precipitation may be difficult using pluviometer data, and since in this work we do not use this data, we simplified the sentence as follows “snow depth/SWE and temperature data” instead of “precipitation data”. In fact, you are correct that also temperature data are used to determine the DD and consequently the ablation state.

This is the in-situ data that we are actually using in this work since they are available in our study areas. However, we discuss the possibility of using precipitation data – that needs to be converted to solid precipitation – in section 2.1.1 also according to another comment raised by the Reviewer later on in this revision.

L168-191: In this section, the authors give a range of possibilities about how an “accumulation” state could be inferred from different data sources. Even though, a closing summary sentence is provided for the section, I still struggle to fully understand the specifics of the method. I suggest restructuring this section to clearly separate i) suggestions about what approaches could be used, and what are the pros and cons of these approaches, and ii) what is specifically used in this study.

Thank you for the comment. We restructured the session as suggested by the Reviewer. Here we report the restructured paragraph:

The *accumulation* identification can be retrieved from a network of automatic weather stations (AWSs) that provide continuous information about the occurrence and elevation of snowfall events, for example, direct SWE measurements or indirect precipitation/SD measurements. Continuous SWE measurements are unfortunately scarcely available. By mean pluviometers and temperature observations, it is possible to split precipitation between liquid and solid (Mair et al., 2013) and identify the state accordingly. However, these stations are rarely installed at high elevations. SD sensors are more suitable for our purpose, but their observations are often affected by wind and gravitational transport, leading to deposition/removal that can be falsely interpreted as *accumulation*/*ablation*. Hence, even if the AWSs are generally situated in locations undisturbed from the wind action, it is more convenient to dispose of a large number of AWSs that need to be filtered to exclude possible sensor errors or wind/gravitational redistribution. In fact, a station is usually representative of a limited area whose extension is highly variable depending on the complexity of the terrain. In the case of a well-monitored area with stations distributed with elevation, it is possible to divide the catchment into different elevation belts where the snowfall events can be considered nearly homogeneous. However, in many basins, this might be far from reality. As a common configuration for snow monitoring, we have a single station located at a high point of the catchment, which is informative enough to identify the accumulation events but not their extent. Furthermore, it has been shown in the literature that estimating the snowfall limit can be very challenging (e.g., Fehlmann et al., 2018).

Therefore, the method can be adapted based on the availability of the in-situ observations. In this study, we made use of SWE and snow depth measurements, since they were available in our analyzed basins. To consider that a snowfall is occurring, the increase in SWE/SD should be greater than a certain threshold SDmin/SWEmin that is fixed at 2 cm for SD according to the values found in the literature (Engel et al., 2017), resulting in a value of 2 mm for SWE when considering the typical density of fresh snow (100 kg/m3). Unfortunately, the basins are poorly monitored in terms of these variables and consequently, the snowfall limit cannot be estimated in an appropriate manner. Hence, we consider that snowfalls occur throughout the snow-covered area of the catchment. We acknowledge that this assumption may result in less accurate SWE estimation, especially in the case of mixed states. For example, snowfall can be observed at high elevations, together with rain-on-snow at low elevations, causing snowmelt. However, we assume that the effect of these events on the total SWE balance is small enough to be considered negligible as we will discuss in Sec. 5.2.

In summary, when the AWSs show an increment greater than a defined threshold, we identify the state as *accumulation.*

L194: DD models are not based on intuition (e.g. Ohmura, 2001) Ohmura, A., 2001: Physical Basis for the Temperature-Based Melt-Index Method. J. Appl. Meteor. Climatol., 40, 753–761, https://doi.org/10.1175/1520-0450(2001)040<0753:PBFTTB>2.0.CO;2.

Thank you, we rephrased with "..which is an empirical model that makes use of air temperature as a proxy of the melting" and added the reference.

L211: it is not clear to the reader why focussing on a "u-shaped" backscattering signal is logical consequence from the methodical issues in forests.

Thank you for noticing this important logical step. We need to clarify that it is still possible to have the "U-shape" in forest. However, the canopy interferes with the signal, and this can result in a "noisier" backscattering signal. A recent work proposed by Darychuk et al., 2023, has proposed an interesting analysis in this sense. They show that the "U-shaped" backscattering signal is more evident in open areas compared to mature forest. They also show that the timing of the runoff onset and duration were less reliable under canopy. However, the intensity of the scattered signal by canopy may depend on the structure, composition, and stem density of forest. Thus, to simplify the processing, we applied the simple rule to consider as valid the runoff onset detected when a clear "U-shape" is visible. We rephrased that part: "As shown by Darychuk et al., 2023, the characteristic "U-shape" signature of the backscattering signal is less evident in mature forest, also depending on canopy structure, composition and density. The signal can be noisier due to the scattered contribution by canopy. Hence, we propose applying the method only for pixels that present a clear "U-shape" backscattering", which can be also present for forested areas."

L290: in line 400 it is stated that the DD factor is held constant during all seasons, please clarify.

Yes, the DD factor is held constant during all seasons. We used the first season to "calibrate" it in the station and then we evaluated the performances for the other seasons. We rephrased the sentence to clarify this point: "It is worth noting that the first year, 2018/19, is also used to set up the DD factor *a* that is then kept constant for the other seasons."

L431 it is tough to see anything in figure 6, regarding the transect locations. Is there any topographic indication why the mid elevation point is so underestimated?

We changed the color of the points of the transect in Fig.6. The purpose of Fig. 8 was to show that also the reference might show an unexpected behavior. In fact, it is true that our proposed SWE at the mid-elevation is probably showing an underestimation, and this is most probably related to a too-late runoff onset estimation. Note that the peak starts decreasing very late if compared both with ASO and the WUS-SR dataset. Further investigations reveal that the mid-elevation point is in a high-density forest area. We will add this consideration to the manuscript. In a previous answer, we were discussing that the runoff onset for forested areas might be less reliable. However, the canopy presence may also lead to overlapping effects as for example canopy interception, differential heat transfers due to the presence of branches in the snow or impact on the surface albedo. This might lead to very different behavior of the snowpack w.r.t. open areas. These overlapping effects are difficult to estimate, and this could be the reason why ASO is also showing an unexpected behavior as we were already addressing in the manuscript. In fact, SWE for the mid-elevation point is unexpectedly much larger than the SWE for the higher point. The presence of forest might affect the quality of the snow depth measurements with lidar or the snow density estimation.

L444: To provide only the very low bias of -5mm in the Abstract I find almost misleading, I must admit. Consider giving more information (eg rmse or r^2).

We agree with the reviewer and the only reason for this decision was to take out from the abstract some of the details. We added also RSME and correlation. Furthermore, we also added the metrics that we obtained when considering the WUS-SR dataset. "It obtained good agreement both when evaluated against HR spatialized reference maps (showing an average bias of -22 mm, a root mean square error (RMSE) of 212 mm, and a correlation of 0.74), against a daily dataset at coarser resolution (showing an average bias of -44 mm, an RMSE of 127 mm, and a correlation of 0.66) and against manual measurements (showing an average bias of -5 mm, an RMSE of 191 mm, and a correlation of 0.35)."

L561: here you state you used SD and SWE data for state indentification. This contrasts the abstract (precipitation)

As explained above, we meant with precipitation data in general what is detected by pluviometers (that however makes the estimation of the new snow more challenging) as well as snow depth/SWE sensors that also return a quantification of the snowfall amount. To make it clearer, we changed this in the abstract with "snow depth/SWE and temperature data" instead of "precipitation data". We also added the temperature as a needed variable for the state determination in the conclusions. These are the variables that we are actually using in this work since they are available in our study areas. However, an interesting future development of this work can be the use of precipitation data – that however needs to be converted to solid precipitation – as discussed in section 2.1.1 also according to your previous comment.

L583: iii) is unclear to me.

We rephrased the sentence in this way: "provide consistent trends when considering the basin scale".

Technical comments

Abstract: make sure no paragraphs are doubled: eg L11-22 and 31-39 should be removed.

Thank you for noticing it. In fact, the paragraph was repeated by mistake.

L83: only true for snow regimes with distinct accumulation and ablation periods.

Thank you. We added this for clarification.

L97: not only to the peak of the accumulation, as you are presenting a method that goes beyond that (and this is still considered reconstruction).

Thank you for noticing it. We removed "up to the peak of the accumulation" to leave it more general.

L100: a general "outperforming" against snow models is a bold statement!

We softened the sentence in this way "The SWE reconstruction approaches showed good performances over large basins and even mountain ranges".

L143: most importantly, the state information is used to add or remove SWE during the reconstruction

This is already written clearly a few sentences later: "The state information is used again in this step i) to redistribute the total amount of SWE calculated for the melting in the accumulation period, and ii) to include late snowfalls that occur after the peak of accumulation to the reconstruction." Hence, we think that we can avoid this repetition.

L152: "Let us introduce": consider change of style.

We changed this with "Three states are introduced…".

L158: which variations? Be more specific.

Thank you for this suggestion. We changed "variations" with "daily increment".

Figure1: "Catchment State" ?

Thank you for noticing it. Accordingly with the revised version of the paper, we will change it with "State".

L159: "melted snow" = melting ?

Thank you, we replaced it.

L486:"..which in turn depends on the quality of air temperature recordings, the interpolation …"

Thank you, we replaced it.

Anonymous Referee#3

I would like to acknowledge the Authors for the great work done in this revised version, which addresses all the reviewers' comments and concerns. I want to especially thank you for the level of detail and care you have included in the point-to-point answers, including those to minor comments in my previous report.

My major concerns have been addressed and the very much improved readability of methods, together with the additional information provided, have facilitated a clear view of the sequential steps and the validity (or not) of the underlying hypothesis. I especially value the so much improved methodological section. Additional information to support different decisions in the approach's building has been satisfactory to understand potential impacts, and the transferability of the results.

We also would like to thank the Reviewer for the effort and the useful comments provided during the review process. We are grateful for his positive feedbacks.

At this point, I just have some minor comments to be assessed in the new version:

1. In their answer to my comment on Section 2.1, the Authors discuss on the assumption of homogeneous occurrence of snow throughout the snow-cover area. From the dispersion graphs newly added, it is easily seen how during 2020 there is an overestimation of the model. From your knowledge of the meteorological data sets during the different years, do you think this is due to such simplification? I am curious about potential differences in the snow cover pattern, and whether that season was more torrential in terms of precipitation.

Thank you for the comment. This is a very interesting comment that is however difficult to be answered without further investigation. In the following, we will try to make some considerations about this concern.
In fact, the season 2020/21 shows an overestimation of the model when comparing the results against ASO. This season was very dry, as you can see from Figure A1. We agree with the Reviewer that possibly we had some snowfalls that happened above the actual snowline elevation. However, from our experience, we have noticed that this kind of error does not so largely affect the results, especially when considering the melting period (as most of the ASO observations). In fact, we remember that the total SWE amount on a single pixel is dependent only on the days that are in "ablation". If we assign more days as the real ones to be in "accumulation", we are simply redistributing the total SWE over more days, thus affecting the actual shape of the SWE trend but not really the peak of SWE. Also, when comparing the results with the WUS-SR dataset, it seems that we are underestimating the most relevant snowfall happening at the end of January. This is for sure a snowfall that involves all the snow-covered parts of the catchment and covers the bare soil, as it is possible to notice in Figure 13c where an important increment in terms of SCA is shown. Hence, we do not directly ascribe the differences w.r.t. ASO to the redistribution of snow throughout the snow-covered area. It is also very difficult to state if the reference is completely reliable given the inverse tendency when comparing our model with WUS-SR. However, we believe that errors can be introduced by the SCA correction in the late season (May-June). As we also discussed in the text (Section 5.2), "an evident case where an overestimation of the SCA is introduced is in the season 2020/21 in May/June for the SFSJR (see Fig.13c). This overestimation is caused by the fact that the AWSs do not indicate an *accumulation* in correspondence with the peaks that occur in the late melting phase. Therefore, the label is corrected according to the majority rule in *ablation* in the case of old snowfall. Most of the pixels were *snow* in previous HR acquisitions, leading to the propagation of the class backward and the consequent overestimation of the SCA. " In summary, we believe that an accurate SCA estimation is very relevant for this work, especially when introducing errors in the date of snow disappearance as shown in our simplified sensitivity analysis. Future work will focus also in this direction.

2. The discussion of the choices made in Section 2.3 is very fine. I still think, however, that the use of DD constitutes a hard limitation for a successful modelling when heterogeneity is large, which, on the other hand, is one compelling fact to obtain new methods that provide SWE on large areas. Figures showing the method performance as related to altitude, slope, and aspect show a relevant shift towards overestimation of SWE in the higher area especially when less snow is present. Do you have some idea on what Is influencing this mostly, error in snowfall threshold or the errors induced by DD methods?

Thank you for this comment. This point can be discussed by looking at Fig. C3. In our opinion, without further analysis, we cannot appreciate a real trend. In fact, we do not really observe a constant overestimation of SWE for higher elevation. From Figure C3, we can appreciate an overestimation, especially for season 2020/21, whose reasons are explained in the previous answer. For the other seasons, we observe a general underestimation of SWE, happening in some cases also when considering high elevations (see for example Fig. C3g). The biggest differences are encountered when considering the slope analysis. As we discussed in Appendix C in L633 (old version): "The slope analysis shows larger differences, especially for some dates (i.e., 9 June 2019 and the three images acquired for year 2020) and when considering steep slopes. The proposed method underestimates SWE w.r.t. ASO. However, we generally expect lower SWE for these steeper slopes that promote gravitational transport. The aspect analysis suggests an underestimation for the north facing slope when comparing our dataset with ASO (except for year 2021)." We are not sure if these differences can be directly ascribed to the use of the DD model, since we would expect an overestimation for the North exposed areas, as the DD method does not consider radiation effects. This is discussed also in Section 4.1 L416 (old version).

Finally, we fully agree with the Reviewer about the limitation of the use of a DD model. In fact, we strongly believe that the future effort should go in the direction of fully exploiting remote sensing data, which has the advantage to be natively spatially distributed observations but which require advanced technical manipulations to be fully exploited in any snow monitoring systems e.g., current satellite missions, even at low resolution, are far to provide daily SCA time series which as discussed in the paper is one of the most important information that we can exploit to reconstruct SWE . Also, we believe that our method could be assimilated into a more reliable physically-based model that better estimates the potential melting and we leave this as future cooperative development.

3. I find the sensitivity analysis at the station site really interesting. I fully understand the limitations explained by the Authors to assess a more complete analysis, and agree that in any case, this adds valuable information to understand the results from the methodological approximation. Maybe some short assessment on the representativeness of this site in both precipitation conditions and the main drivers of snow dynamics could further facilitate the interpretation of the whole set of results.

Thank you for your comment. The Volcanic Knob station is located at an elevation of around 3050 m. Being the average elevation of the catchment around 3070 m, the station is thus well representing the mean behavior of the catchment. We will add this relevant information to the manuscript L321 (new version). Given the presence of relatively long-term records for the VLC station, we have presented in Appendix A – Fig. A1 – the climatology of SWE for the analyzed years. This gives an overview of the precipitation conditions at the site. We also would like to highlight that the San Joaquin basin is a snow-dominated catchment where high-elevation snowmelt strongly contributes to runoff during spring and summer months. The same is valid also for the Italian basin presented in the manuscript. However, to fully confirm the results of the simplified sensitivity analysis, it would make sense to extend it to more stations where snow pillows are available.

4. With the new additions and graphs, I'm not really sure that Figure 11 adds relevant information, and I would consider eliminating this as the hydrological balance is not the goal of the manuscript, nor is it treated in detail and completeness as the other results included in this final version. But this is just a minor reflection from my side, up to you to decide.

Thank you for your suggestion. We agree with the Reviewer that Fig. 11 is not completely representative and also that the hydrological balance is not the purpose of the paper. However, we would prefer to keep it since we believe it represents an interesting qualitative result for this second catchment where we unfortunately do not have a proper spatialized SWE reference. Furthermore, we believe that it can stimulate further research.

Thank you again for the work done and your consideration addressing all the answers. A very good work, indeed.

Thank you for your contribution and the many constructive feedbacks.