Cover letter

Dear Guillaume Chambon,

We want to thank the reviewers for carefully checking the revised manuscript. We addressed all comments, which helped to further improve the manuscript and remove inconsistencies.

Below you can find our answers to the reviewers comments.

Looking forward to your evaluation,

Best,

Michael Matiu

Referee 1: Ross Brown

Many thanks to the authors for their diligent and comprehensive responses to my and the other reviewers' comments. The manuscript is significantly improved and I recommend the paper for publication after the authors have attended to the following minor comments:

Thank you for the positive feedback and for carefully checking the revised manuscript.

1. The regression analysis contains a couple of loose ends, namely the use of different methods (OLR and GLR) for different variables, and the use of bounded definitions for SCD that the authors indicate has some influence on trend results at low and high elevations. The use of a standard non-parametric trend analysis method such as Mann-Kendall would solve the first issue, while less restrictive definitions of SCD would solve the latter e.g. SCD computed over entire snow year and separately for the first and second halves of the snow year.

Thank you for pointing out these aspects. Concerning the SCD issue, indeed we already included during the first revision of the manuscript separate calculations for three different periods: first half was November to February, second half was March to May, and the full period November to May. However, we noted that SCD splitting does not solve the bounded nature of the metric. In the case of the higher elevation sites, the issue could be solved by increasing the length of the snow season to the full year (as we only used November to May), thus capturing the start and end of the snow cover. However, not all station records covered the entire year.

Instead, the solution we identified is to include in the analysis a probability distribution that is suitable for count data and that accounts for overdispersion, such as Negative Binomial. This is also able to handle the many zeros of SCD at lower elevation. We then compared the results of the Negative Binomial model to the other trend models (see below).

Since you mentioned the non-parametric MK test (with Theil-Sen slope estimates, we presume), we systematically evaluated the influence of the trend model on the results. We compared the estimates of GLS to OLS, to non-parametric MK with Theil-Sen, and for the SCD variables also to a Negative Binomial model. Both trend magnitudes and significance values do not differ substantially by choice of trend model. Trend magnitudes are highly correlated (on average > 0.95), and the trend significance agrees on average in 88% of the cases. We added details of all comparisons in the supplementary material as well as a short summary of these new results in the manuscript at Section 2.5 (methods: trend analysis).

Given these results, we decided to consistently use the GLS model throughout the manuscript and remove OLS.

2. Figure 8 is a useful addition, but did you consider computing regional series as standardized anomalies? This would highlight common responses and may allow the construction of a pan-regional series for each elevation range.

We created the suggested figure and provided it in the appendix. It seems that the SCD anomalies are highly correlated within regions, so a pan-regional series might work there. However, for the snow depth indices, this is not the case. Consequently, we were reluctant to create pan-regional series at this stage.

But as you implied, the anomalies highlight the common responses (or deviations) better, so we think it's a useful addition. We also added a sentence on this in the conclusion.

3. I flagged two places where I think some improvements could be made in the wording:

Section 3.7: the material on the homogenization of the series would fit better with the discussion of data homogeneity starting in line 180.

We moved this paragraph as suggested.

Line 523-524: Suggest removing the sentence "Such low R2 values…" as it does not add anything useful.

Done.

4. The fact that trend significance increases with elevation and is stronger in the spring period are important points that you should consider highlighting in the conclusions and abstract, as they are consistent with enhanced albedo feedbacks in mountain regions. I'm currently unable to access my collection of snow literature, but I recall there being several papers published on elevation-dependent albedo feedbacks that you can cite as part of this discussion.

Thank you for this suggestion. We highlighted this aspect in the abstract and conclusion, also citing relevant literature.

Referee 2: Anonymous

I would like to thank the authors for the additional effort they put in the revision of the manuscript. Readability and completeness were much improved. I particularly like section A.3 which is now very detailed and comprehensible. I have only some minor additional comments (the page and line number refer to the latest manuscript version with track changes).

We want to thank you for carefully evaluating the revised manuscript and for your valuable comments, which helped to identify many inconsistencies.

Point comments

L55: I find "after accounting for elevation" a bit odd. Do you mean trends of the same elevation bin differ amongst regions?

Exactly. We modified this to "at the same elevation".

L145-146: I don't understand this sentence. Why is synthesizing studies into a common Alpine view relevant for providing snow cover information at the regional scale?

Sorry for the confusion. Regional is somewhat ambiguous, and we meant the mountain range or larger extents, so we modified the sentence accordingly.

L331-332: I'm not sure if I understand this correctly: you tested the clustering with 2 – 8 PCs, right? It's a bit confusing because a couple of lines above, you write that you only retain the first five PCs.

Thank you for spotting this inconsistency. We forgot to remove the sentence in the previous lines during the revision.

L373-374: Why do you only filter time series based on the months April and May?

We only removed the respective months and not the other months. We made this clearer in the manuscript.

L390-391: You state that GLS performed better than OLS (but only for a small model fraction). But this would already justify the application of GLS – or not?

We followed a parsimonious approach, where we tried to keep the simpler models. However, also the other referee mentioned this inconsistency. We systematically evaluated differences between various trend models, and since they were marginal, we opted for GLS model for all analyses. We added explanations and results of the trend model comparison in the methods and supplement.

L540: Trends were not discussed yet à I would rephrase the beginning of this sentence

True, an artifact of the manuscript restructuring. We changed the linking sentence to match the previous section.

L547: The term "station" might be ambiguous here. It refers to snow stations, right? And not meteorological stations that measure e.g. air temperature and precipitation.

True. We changed "station" to "site" in this section, when it refers to temperature and precipitation.

L568-571: I'm still struggling with this sentence. Also because trends have not been discussed.

Yes, this does not make sense at this point. We removed the part from this section, and added instead a sentence in Sec. 3.7 (Outlook), where we already have discussed detection and attribution.

L701-703: Do you have an idea what could cause this difference?

We tried to reproduce the results from Klein et al 2016, and found the following: For maxHS, the differences are caused by the different periods (1970-2015 vs 1971-2019). For the SCD variable, this is also because of the different period, and in addition also because of the different season definition. We used November to May, while Klein et al looked at the whole year (this affects especially the higher elevation sites). We added some sentences on this also in the manuscript, since this gives some evidence to our introductory statement why the synthesis is challenging.

L704-705: I'm confused by this sentence. Do you mean that in terms of absolute values, your study and the one from Bach et al. (2018) do not agree?

Exactly. We tried to make this clearer.

L718: I would replace "spatial and elevation" by "horizontal and elevation"

Done.

L727: I would be careful with the statement "their significance is limited for hydrological applications" because higher elevations typically store more snow than lower elevations (hence the 0.7% can be misleading)

True. We modified the statement to: "While the elevations above 3000 m only cover a minimal area (0.7% of the area studied here, see Fig. 2(c)), they store large amounts of snow: Figure 6(a) gives an indication of the expected increase in HS with elevation. Long-term monitoring is extremely challenging at elevations above 3000 m, and the snow cover at these elevations is relevant for hydrology, mountain ecosystems, glacier dynamics and mountain (ski) tourism. "

L794: Are depth of snowfall (HN) not used at all in this study? Somehow, I had the feeling they were used for quality control of the data. Or were they only processed for the harmonized data set?

Thanks for spotting this. Yes, we used HN to same extent in the quality checking, but have not analysed it all otherwise. We made this clearer.

L892: I guess some horizontal distance was used to select appropriate stations, right?

5

For this step of the quality check, no. Given the climatological nature of the screening and the stronger dependency on elevation, we found horizontal distance here less important. We tried to make this clearer in the manuscript.

L965: What does "halving distance 250 m" mean?

The weights (for the average) are based on an exponential decay, where the weights are halved every 250m. This is a more interpretable transformation of the decay parameter (lambda). We added more clarification in the manuscript.

L994: What does "non-zero true values" mean?

This was a bad choice of formulation. We meant the non-zero held-out values. This has been changed in the manuscript.

L1000: This sentence is oddly stated: How can the gap filling be unbiased with an overall non-zero daily bias?

We changed "is unbiased" to "has extremely little bias".

Stylistic comments and typos

L228: "merged to" à "merged with"?

Done.

L235: "used for the" à "used from the"

Done.

L283: "some sources ended some date in between" à "some sourced ended in between"?

Done.

L360: "made sense" à "are meaningful"

Done.

L491-493: This sentence is still a bit hard to read. Maybe one could write: "and see if an additional third cluster would emerge (as in the north)…"

Done.

L770: "gridded area" à "gridded product"

Done.

L786: remove "one of those" or write "SCD and maxHS are amongst indicators least affected"

Done the latter.

L976: "Visualization of some steps of the gap filling algorithm."

Done.

L1226-1227: change to "which includes scripts for the following tasks: reading in different data sources, performing all data pre-processing, quality checking, gap filling and statistical analysis."

Done.

Figures and tables

Figure B1: The figure caption is maybe missing

The version with track changes looks so, but in the clean version it is there.

Table B2 à caption: correct "Empty cells no stations with…"

Done.