Dear reviewer, thank you for your comments and suggestions. Please find our replies below in blue.

The authors have performed homogeneity testing on snow series using three different methods on two different snow variables. The paper is well written with good structure and nicely presented results. It is a good addition to the scientific literature on homogenisation of snow depth series. However, I do think it would benefit the paper to include some results that shows the impacts of the homogenisation (difference between raw and homogenised series, trends…).

We understand the interest for trends and differences between raw and homogenised data and agree that such an analysis is relevant and important. Our reasons for not including these analyses in the current study are three-fold:
- In order to obtain homogenised data, the series have to be adjusted at the break points. The adjustment process itself is far from straightforward (and so far, not really looked at for snow) and including it in this study would not have done it justice.
- We considered the topic too important and complicated for a mere paragraph, so we decided to give it its own dedicated article, which is currently being prepared. In undertaking this process, we are currently experiencing large method dependent differences, which merit further investigation and more detailed reporting to share with the community.
- Including trends or differences of homogenised series would also have shifted the focus away from the break point detection problem.

I agree with reviewer 1 about it not being clear why break detection depends on elevation. I suggest adding a version of what you answered reviewer 1 to ch. 4.1.3 or another suiting place in the paper to explain your motivation for looking into this.

Added information to 4.1.3 and rephrased paragraph to:
As the availability (and quality) of suitable reference series for each candidate station is key for a proper break point detection and our stations range from 200 to 2500 m a.s.l. To test the hypothesis that lower stations might not have enough suitable reference series for proper break point detections a possible elevation dependence is investigated. To do so, a possible altitudinal or amount-of-snow influence on the break detection capability of the methods, the break [....]

Fig. 8 could use some refinement:

What I understand from the figure text is that valid breakpoints detected by two methods are shown in grey but breaks only detected by one method are also shown in grey. This is confusing. And are breaks detected by only one method shown in the figure?
Thanks, the figure caption was wrong, only valid breaks are shown.
Comparison of valid break points found for HSavg and dHS1. Valid break points, detected by at least two methods, are coloured grey (dHS1), yellow (HSavg), or blue (both). The shape indicates which method detected the break and the size is the corresponding break magnitude correction.

- In the legend I suggest using a color other than grey for the circle, square and rectangle (for acmant, climatol and homer) or another color than grey for dHS1, whichever is easiest.
  Colours removed from the shape symbols (acmant, climatol, homer) to better distinguish between methods and variables.
- In addition, there is a “NA” in the legend that looks out of place.
  Yes, done.

Please see the attached pdf for the rest of my mostly technical and minor comments.

Comments from the PDF

L7: 45 breaks in how many of the 184 series? (i.e. please add some information about how many of the 184 series were classified as inhomogenous)
Sentence rephrased to: 45 valid break points in 41 of the 184 series investigated.

L10: done

L20: done

L83: done

L120: done

L125: Should there be a "plusminus"-sign in front of 2?
Yes, rectified

Table 1:
Done

173: In this paragraph, could you also please mention the complementary use of the two variables HSavg and dHS1 and what you want to evaluate by using the complementary approach?
Sentence added: The main focus is on breaks in HSavg, however, the opportunity to use dHS1 alongside as a complementary break point detection approach is discussed in Section 4.3.

192: Please rephrase/clarify this sentence. I read this and thought it was not consistent with what Fig. 3A shows (where most Acmant break corrections are between 10-19 %, but where this is not the case for Homer and Climatol). I was confused by the last part of the sentence ("10-19% for all three methods").
Sentence rephrased to: 10-19% (ACMANT and Climatol) and 20-29% for HOMER.

Figure 3: delete text
Done

L210: done

L213: done

Figure 8:
See comments above

Table 3: Please briefly explain the "code" and "combination" column in the table text.
Changed code to Name and added combination information in the caption. Rephrased to:
Table 3: List of valid break points concordantly identified in 184 investigated Swiss snow time series. Stations are ordered according to break year
and series with break points identified in both HSavg and dHS1 are marked bold. Combination code refers to ACMANT (A), Climatol (C), and HOMER (H).

L261: Instead of the dash, should it be plusminus or ~? Also, add space beween 2 years.
Yes, rectified

L317: changed to:
the six break points identified in both HSavg and dHS1 reveals

L331: rephrased sentence to:
[..] based on a combination of reasons.

L386: 25 % of what? Please clarify.
Rephrased in relation to the total number of series investigated:
We identified 45 valid break points in 41 of 184 (22%) series investigated using a complementary approach [...]