Dear referee

Thank you for your comments and suggestions. Please find our replies below in blue.

The paper is well written. The presentation is clear and well structured. It has been a pleasure to read it.

The paper presents a study assessing the performance of three different applications for identifying homogeneity breaks in snow depth time series. The analysis and results are well described and discussed. My main concern, which is not very large, is related to the comparison of the applications. There are obviously several subjective choices to take in applying both Homer and Climatol, and less in Acmant. One issue the authors mention, and which I recommend to pursue more, is the role of the reference network and how they are established, and how that influences the results. I have a feeling that this is as well an important issue as the break detection algorithm itself. In order to have a fair, consistent comparison of the break detection capabilities I would therefore challenge the authors to apply the three methods with the same reference networks. Would it be possible to apply them on all three combinations of networks?

We agree that selecting/building suitable reference series is as important as the mere detection of change points itself. However, the network-selection process is an intrinsic component of ACMANT, Climatol, and HOMER and cannot be separated.

In ACMANT and Climatol the options for the user specifying the reference (sub) networks and their selection basis are limited. In addition, Climatol uses a geographical distance criterion as the default to select localised reference networks, whereas ACMANT uses a pre-set correlation threshold of 0.4. While it is true that HOMER does allow a geographical distance selection option there is no easy way to match these to the Climatol selections, as HOMER does not consider the vertical distances.
We discuss the different correlation thresholds we applied in HOMER (0.8) and ACMANT (0.4) and the possible implications of this in Section 5.1.

As it is not possible to exclude the network-building from these methods and simple run the break point detections alone, we choose the settings to be as realistic as possible for each method. However, as we are looking for a practical solution and comparison, (and accepting the network-building as an intrinsic component of the method) running the three methods with the same input data is as close to fairness as feasible.

Minor comments:

Line 27: Typo: contemporay → contemporary.
Done

Figure 7: Difficult to read. Use a lighter shade for the terrain, and make the black dots smaller. Consider dark-grey contour lines (rivers, borders).
Changed the contrast to increase legibility