Cover letter

Dear Guillaume Chambon,

Thank you for the opportunity to submit a revised version of our manuscript. We are very grateful to all reviewers, whose comments substantially improved our manuscript. We responded to all questions in the interactive discussion and put the responses also below for convenience.

Here a brief summary of all the changes:

- Literature summarized in table with main elements only in the introduction section
- Removed the moving window analyses
- Added a section on inter-annual variability
- Added analyses of seasonal snow depth and snow cover duration indices
- Changed our modelling framework to account for changes in variability (as consequence of the inter-annual variability analysis)
- Restructured auxiliary material (previously appendix and at repository; now appendix, supplementary material, and repository)

More details can be found in the revised manuscript. The version with track-changes is rather noisy because we re-read the whole manuscript and also changed the order of sections to be more consistent (30-year versus 49-year analyses).

Thanks again, and we wish you happy and safe holidays and a great start in the next year,

Michael Matiu

Reviewer 1

An extensive set of snow depth measurements from the European Alps is evaluated to explore snow climatology regions and temporal trends. Raw data is collected from various sources and subsequently harmonized, quality controlled and temporal gaps were filled. Based on the snow depth time series, five regions with different snow climates are derived using Principal Component Analysis and k-means clustering. Linear long-term trends and short-term trend variability are computed from the snow depth time series for the five derived regions. Finally, snow depth observations are compared with gridded air temperature and precipitation data, which is either reanalysis based or inferred from spatially interpolating observational data. The authors find decreasing trends of snow depth for the majority of the stations and substantial differences in trends between regions.

This study addressed a very relevant topic: the evolution of snow depth in the European Alps during the last half century. The authors quantify these changes based on an extensive compilation of in-situ time series of snow depth from different regions. Additionally, the authors provide the unified dataset as an online resource, which can be very useful for further applications. The manuscript is well written and applied data and methods are sufficiently explained. My comments concern therefore primarily smaller ambiguities and (potential) errors.

We thank you for the detailed review of our manuscript, the appreciation of our work, and the constructive comments, which we shall address in a revised version.

General/major comments

• Some sentences are extremely long and should be shortened to increase readability.

We shortened sentences as much as possible throughout the manuscript.

• **Selection of different time periods**

I'm confused about the different time periods that were used for this study: The PCA and k-means clustering was performed for 1981 – 2010, long-term trends were computed for 1971 – 2019 and shortterm trends for 1961 – 2019. Could you elaborate in more detail why you use (these) three periods?

We chose different periods because of the different nature of the analysis. The PCA and k-means aimed at having the largest spatial density, so we chose the 30-year period with the largest number of stations. The long-term trend analysis aimed at a tradeoff between coverage of stations and as long as possible period. The moving-window short-term trend analysis will be removed in the revision (see responses to Ross Brown's review comments). We shall add an explanation also in the manuscript and provide more information in the data overview sections on these two periods and station subsets.

• **Which data (raw, gap-filled) was used for which analysis/plot?**

In the manuscript, I was sometimes a bit confused which data was used for which application. E.g. in section 2.4 you use the raw data without any gap filling, right? But for the subsequent analyses you always use the gap-filled data (as explained in appendix section A.3)? This point should be more explicitly stated in my opinion.

Thanks for pointing out this ambiguity. Actually, we used the gap filled data for all analyses. We now explicitly state this in all related method sections (overview as well as statistical analyses).

• **Gap filling method**

I struggled to follow the explanation of the gap filling method – particularly from line 641 to 654:

o I do not understand what "crossing a calendar day window with a year window" means.

o Shouldn't the "window of 31 days" be "window of gap length + 30 days"?

o "mean of the daily values" -> does this refer to the daily climatology?

o "and the weights were based on the vertical distance between candidate and reference station." àwhy are horizonal distances not considered?

o Finally, I wondered if reconstructed values have a smaller temporal variability (on smaller scales) because you apply climatological values in your method. But this is not the case, right? Because you only compute the ratios from climatological values (daily means)?

We rewrote the description of the gap filling in an algorithmic way and also provide an additional explanatory figure. We hope that this helps to clarify our procedure. Regarding your specific comments:

- The horizontal distance was only considered for selecting candidate stations and not for the weighting, because we wanted to have univariate weights and not multivariate. In that case, we found the vertical distance to be more important than the horizontal distance.
- The temporal variability of the reconstructed series should not be affected. The climatological values are only used to calculate the ratios. The reconstructed values is then based on the daily value(s) of the neighbouring series. Consequently, the daily variability in the reconstructed series stems from the daily variability in the reference series'. The only loss of variability could occur because the reconstructed value is an average of up to 5 values from up to 5 reference stations.

Point comments

L150-151: I don't understand this sentence – could you rephrase it?

Done.

L157-158: this sentence is a bit odd: why do three different climate forcing zones create four main climate regions?

Yes, it is counterintuitive. However, the three forcings combined with the topographic effects result in gradients along the North-South and East-West directions. And these two gradients, if intersected, result in four regions. We modified the wording accordingly.

L179-189: I struggled to read this sentence (because it is so long). To increase readability, the providers from Italy could e.g. be listed with bullet points.

As suggested, we restructured the data providers of Italy as bullet points.

L195-196: "with a few expectations of monthly/seasonal data form the HZB and SMI." -> was the monthly/seasonal data also used in this study?

Actually, no. We removed this part of the sentence, to avoid misunderstandings.

L196-199: I'm confused by this sentence: automatic measurements are used both from France and the Aosta Valley, right? Then I would change "only for France or" to "only for France and". Anyway, I think the sentence should be rephrased to improve comprehensibility.

The sentence was split and rephrased.

L205: "(see also Fig. 2b)" should be "(see also Fig. 2c)", right?

True. Thanks for spotting the error.

L245: why is the data for Austria only available until 2016?

Because of the processing and quality checking performed by the Austrian Hydrographical Service, it takes some time until the data are published online (we added this information also in the revised manuscript). In our case, when we accessed the data (early 2020), only data until 2016 were available. Since a few months, the records have been updated to include 2017, however, we cannot manage to re-analyze this update in the revision and, furthermore, results are not expected to change significantly including a one year extension of the Austrian series.

L253: "The frequency by elevation (50 m bins were used to calculate to proportion)…" -> I don't understand this part.

We plotted so-called frequency polygons, which are basically histograms shown as lines and not bars. This makes it easier to compare different distributions. But, as for histograms, it is necessary to specify the bin width. We tried to improve the caption to make it more understandable.

L257-258: why is this criterion applied? Wouldn't e.g. a threshold of 50% make more sense?

Yes, a different threshold would also make sense. However, then, also the meaning of the figure would be altered. We chose this simple threshold to show availability of stations per year. Another option would be to show stations with the threshold you proposed, but then it also depends on which season the 50% apply (Dec-Feb, Nov-May, or Oct-Sep). Yet another option would have been to show only the stations used in the analysis. We decided to stick to our simple threshold, and tried to clarify the intent.

L273: did you consider to use the "elbow method" to find the optimal number of clusters?

We shall look at the elbow method for the revision. We also compared it to the average silhouette coefficients, which we used initially. The elbow method identifies 4 clusters as optimal if we apply the k-means on the scaled observations, and 5 clusters if we apply it on the PC matrix. The silhouette analysis identifies 4 clusters as optimal, if applied on the scaled observations. If applied on the PC matrix, then 2 clusters are optimal, followed closely by 5 clusters. We checked maps of the clustering results for all combinations (see figure below; we also shall add it in the supplementary material), and, to be honest, all choices of clusters make "sense". The different number of clusters and whether observations or PC results are the input, all highlight different aspects of the snow depths and their hierarchy (e.g. as seen by increasing the number of clusters). These are the elevation, North-South gradient, and East-West gradient. We finally decided to leave our analysis as it was with 5 clusters, because they agree best with our knowledge of the climatic and topographical drivers of snow depth in the Alps. This information on this process shall also be added in the revised methods section.
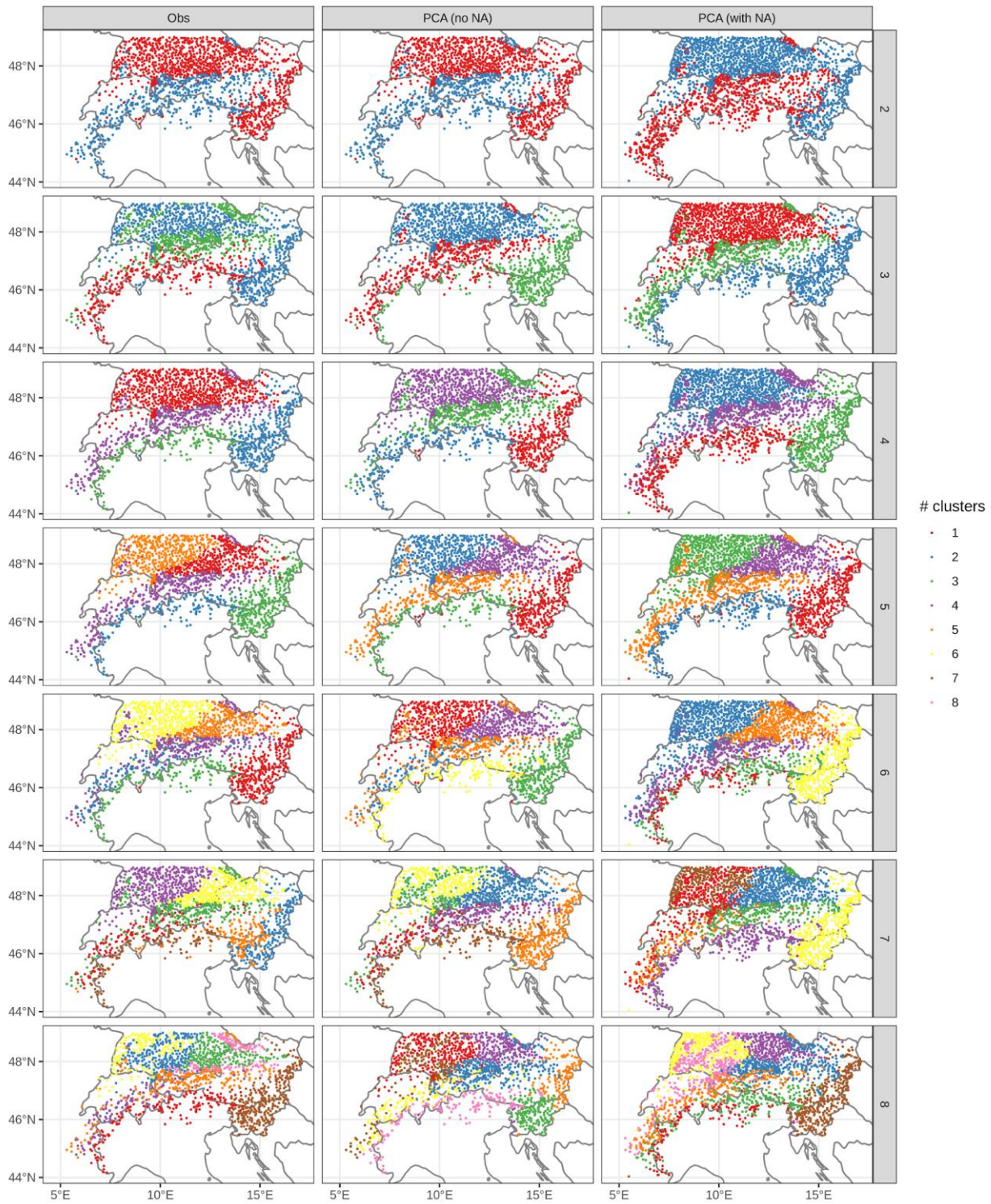
Figure R1: Results of k-means clustering. Rows show the number of clusters (order is arbitrary, so colors might not match within a row). Columns identify the input matrix for the clustering algorithm: For Obs scaled daily observations of snow depth were used and for PCA the PCA matrix. The two PCA columns stand for the standard PCA that does not allow missing values (no NA), which corresponds to the same station set as in Obs, and the second PCA column is for the modified PCA algorithm that allows missing values (with NA) and has a higher station coverage.

130

L273-274: what do you mean by "as well as clustering directly on the daily observations."?

135 We applied the k-means clustering on the PCA matrix with the principal components as input, and, as comparison, we also applied the k-means clustering directly on a (scaled) matrix of daily snow depth observation. We tried to make this clearer in the manuscript.

L287-288: remove "using assessed"

140 Done.

L337: "and is highly correlated to elevation up to 1000-1500 m, and mostly constant above." I don't understand this part.

What we meant is that the PC2 is related to elevations up to 1000-1500m. We removed the word "correlated" and rewrote the sentences to make it clearer that PC1 and PC2 are both elevation driven, but PC1 mostly >1000m and PC2 mostly <1000m..

145 **L348:** the clustering was performed on the 5-dimensional PCA loadings, right?

Yes. We rewrote the sentence to make it clearer.

L367-368: please provide a reference here for the HISTALP subregions

Done.

L370: what you do mean by "estimated data-driven"?

150 We meant that the clustering was performed automatically using snow depth data, and no manual re-assignment or modification was done on the clustering results. We modified the sentence in the manuscript accordingly.

L377: what do you mean by "unique stations"?

Unique in the sense that they have no similar neighbours in any cluster. We now added this information in the manuscript.

L386: I would not write "matches" here because there are 4 vs. 5 regions. I would rather write that the obtained regions are

155 similar.

Rewrote according to your suggestion.

L404: does this statement refer to a specific month? Or the entire winter?

To the whole winter. Clarified it also in the manuscript.

L408-409: "mean North & high Alpine" is odd. Do you mean: "While in December, the mean negative trend was stronger in

160 North & high Alpine"?

Thanks for pointing this out. We rewrote the sentence.

L422: "Points with lines indicate" -> "Lines indicate" (or are there points with no lines?)

No, you are right, all points have lines, even though they might be hardly visible. We rewrote the legend.

L467-468: "Moreover, we assume that most of this seasonal imbalance is because there is no or no significant snow cover in that month" I don't understand this part

We meant the seasonal imbalance of station observations, since some low elevation stations do not record outside the winter season. We rewrote the sentence accordingly.

L493: I guess "~100m" should be "~100 mm"

Yes, thank you. We also changed it to the ≈ sign.

L515-516: "because this implies less chances that precipitation falls at the "right" time." I don't understand this part

We replaced "right time" with "concurrent with low temperatures".

L532-533: I'm not sure if I understand this sentence correctly. Do you mean homogenization is not so important because such a large number of stations is used?

Partly. Homogenization is important. But given the extent of our dataset it was impossible to apply a common framework to all the data. We made this clearer in the manuscript.

L589: how do you define a network? A country?

By data source i.e. data provider. Clarified it in the manuscript.

L612: how is this surrounding band defined in terms of horizontal distance?

Horizontal distance was not considered here. We added this in the manuscript.

L673-674: I don't' understand this sentence.

We assume you refer to the explanation why our relative MAE is not a "true" MAE. In the standard way, the relative MAE is defined as $\frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - x_i}{x_i}\right|$, while our modification is $\frac{1}{n}\sum_{i=1}^{n} |y_i - x_i|/|\underline{x}|$, where $\underline{x}$ is the average of all x_i. We added the formulas also in the manuscript to make it clearer.

L686: what is meant by "ablation scheme of the different stations"?

The local climatic and topographic characteristics that influence ablation. We modified the sentence accordingly.

L701-702: "gap filling snow depth series using simulations of the Crocus snow model for the French Alps" -> I'm a bit confused by this part. Does it state that gap filling was performed by running the Crocus snow model with meteorological forcing?

Yes. Actually, the Crocus simulations were performed independently of the gap-filling used for this study, but we found it interesting to compare the two approaches, since they were both available at the same sites. We added some more information on this in the manuscript.

L721: what is meant by "original observations"? Available observations?

The observations available before gap filling. Modified the sentence to make it clearer.

Figure 1b: this panel is hard to read. Could you enlarge it? To increase readability, station density could be plotted instead (i.e. the number of stations per a certain area).

We splitted the figure in sub panels (for available and used), and put the station density underneath the points for a 0.5*0.25 deg grid. We still show points, because we think they are important.

Figure 2b: how was the polygon for the DEM generated? With a convex hull?

No, a manual outlining along the stations, because a convex hull would include most of the Po valley in Italy, for which we do not have any stations.. We clarified this in the caption.

Table 2: there are typos in the first row (e.g. "(0,1000] m").

We modified the elevation intervals to say "Elevation: (0,1000] m" etc.

Table 3: the spacing between the columns should be improved (-> it is currently confusing that the columns "DJF #" and "MAM mean (min, max)" are so closed together)

We aligned all numeric columns to the right, so the spacing should be better now.

Figure A1: I guess a subset of stations was used to produce this figure, right?

Yes, it is explained in the text, and we added this information to the figure caption.

Figure A2: which statistical quantities (percentiles, outliers, etc.) do the points, lines and box edges represent?

We added this information in the figure caption.

Figure B4: how is the numerical quantity "silhouette width" computed?

With silhouette width we mean the silhouette value or coefficients, sometimes also called width. We modified the figure and provide the formulas for how the silhouette is calculated in the methods section.

Figure C4: For completeness, the table for MESCAN-SURFEX (March to May) should also be shown

Added.

Stylistic comments and typos

L120: "1960–1990. (Lejeune et al., 2019)." -> "1960–1990 (Lejeune et al., 2019)."

Modified.

L144: "while Section 4 concludes." -> "while Section 4 convers conclusions."

Modified to "Section 4 provides conclusions".

L147: "with their typical arc-shaped" -> "with their arc-shaped"

Modified.

L231-232: change to "Station numbers are shown for fresh snow (HN) and snow depth (HS) time series."

Modified.

L279-280: I would remove the line break here.

Thanks, but since we removed the moving window analysis in the revision, this comment resolved itself.

L312: "significantly to the" -> "significantly from the"

Modified.

L348: "There were" -> "This yielded"

Modified.

230 **L353-354:** "South of the main ridge, there were two regions:" -> "Two regions emerge south of the main ridge:"

Modified.

L362: "as has in the north" -> "analogue to the north"

Modified.

L376: "the station in common, and the same common stations" -> "the stations, and the same stations"

235 Modified.

L400: I would rather use present tense here (and in the following lines).

We noticed our inconsistent use of present and past tense in the results. We mostly prefer past tense for results, and adapt the complete results section accordingly.

L469: remove "supposed to be"

240 Modified.

L561: there seems to be a space in the word "scientific"

Yes, it looks like this in the PDF of the paper, but in the Word version everything is fine.

L661: "were useful" -> "are justified"

Modified.

245 **L664:** "has not been yet used" -> "has not yet been used

Modified.

Reviewer 2

General comments:

250 This paper provides an analysis of snow cover regional variability and trends over the European Alps based on a new in situ daily snow depth dataset developed through the collaboration of more than 20 institutions from six countries. The dataset covers the entire European Alps with more than 2000 surface snow depth observations, and represents an important contribution for research and development. The authors are to be congratulated for their efforts to develop this dataset and in particular, to make it available to the research community. The
255 creation of a pan-Italian snow depth dataset from various agencies is a noteworthy achievement.

The paper presents the results of a PC and cluster analysis to characterize the regional snow climate, along with trend analysis to document trends by climate region and elevation over almost 50 years (1971-2019). The paper is in general well-written and clearly explained, and is close to publishable quality once some issues with overly-long sentence construction are rectified. I have three main comments concerning the methodology. First
260 (comment #6 below), I question the need for the moving window analysis for trend variability, and recommend it be removed from the paper. Second (comment #10), the PC results reflect an uneven spatial distribution of stations with oversampling of elevations below 1000 m and undersampling of elevations above 2000 m. It is unclear to what extent this distorts the analysis results compared to those obtained based on a gridded representation of the station data that evenly samples the full spatial and elevation domain. Third, the paper
265 provides no insights into interannual variability of snow cover and its relative magnitude compared to the long-term trend. The authors may feel this is beyond the scope of the current paper, but presenting trends without discussing the interannual and multi-decadal variability is a major oversight in my opinion.

I look forward to seeing the revised paper and congratulate the authors again for their significant contribution.

Ross D. Brown, Canada (ross.brown@canada.ca)

270 We thank Ross D. Brown for the in-depth review of our manuscript, his positive appreciation of our work as well as the detailed and constructive suggestions for improvements, which have been very helpful in revising the manuscript. In order to address the 3 major comments, in the revised manuscript we plan to introduce the following modifications:

1) The moving window analysis will be removed.
2) The uneven distribution of stations across elevation is an issue. However, we think that interpolating the stations
275 to a grid could introduce more uncertainty, given the complex topography in the European Alps. While it is certainly a good exercise, it goes beyond the scope of this paper. See also the detailed comments below for more on this issue

Detailed comments:

1. Lines 74-75: Suggest rewording as "The main limitation … that their number decreases sharply with elevation, with few stations available above 3000 m in the European Alps."

2. Lines 90-126: There is a lot of useful material presented here on published snow cover trends in the various countries, but it is difficult to read with very long sentences joined with rather unwieldy constructions like "which, however". I recommend you organize this material in a summary table, and provide a few lines that capture the common elements. This would lead very nicely into the paragraph starting on line 128.

3. Lines 155-160: It would be instructive to show the main climatic divides on Figure 1.

4. Lines 204-206: consider rewording as "Many stations contain an important data gap for the 1981–1997 period that rendered a large fraction of the stations unusable for this study."

5. Line 279: Suggest deleting the following "The predicted variable was the mean monthly HS and the only predictor the year (shifted to 0)" and replacing the previous sentence with "Linear trends in monthly mean HS were computed separately for each month from November to May for stations with complete data in the period."

6. Line 280: "The second approach was a moving window approach that aims at identifying the short-term changes in trends." I think it would be clearer to say "A second moving window approach was used to examine the variability in 30-year trends over the period from 1961". I would consider removing this analysis for the following reasons: (1) the lack of a clear rationale for the analysis, (2) the inconsistency introduced by the

different start period (1961 vs 1971), (3) the fact that overlapping windows are not independent, (4) the use of what is essentially an arbitrary 30-year period for the trend, and (5) the fact that the network density changes over time. I think it would be more instructive to look at the signal-to-noise properties of the 1971-2019 trend, by breaking it up into the amount of variance explained by the trend vs the amount of variance explained by interannual variability. Mapping the two quantities would highlight areas where trend was stronger relative to natural variability and vice versa.

Thanks a lot, we clearly see your point and agree to a large extent. We therefore removed the moving window approach from the manuscript. Instead we introduced the analysis of interannual variability (see also your comment no. 11 below). This analysis presents the time series plots of the mean monthly HS (averaged over 500 m elevation bands because of the number of stations) and a short discussion of the fraction of variance explained by the trend.

Looking at the time series figures, we noticed some changes in the interannual variability across time. The most prominent is the decline in variability at the end of the season associated with the decrease in mean HS, which approaches zero. We decided to account for these changes in the variability in our linear model by including a time coefficient for the error variance (=interannual variability). This results in replacing the standard OLS model with a GLS (generalized least squares) model. For our regression formula $y_t = \beta_0 + \beta_1 t + \epsilon_t$, OLS has a constant error variance $Var(\epsilon_t) = \sigma^2$, while with a GLS we can allow the error variance to depend on time: $Var(\epsilon_t) = \sigma^2 * exp(2 * \gamma * t)$, where γ is another estimated coefficient that indicates the change in variance associated to t.

The trends in mean HS are not affected, but, with the new model structure, we were able to account for changes in the variance. More details on the model specifics can be found in the revised method section. We shall also discuss the results in the new section on interannual variability.

7. Line 297: Not clear what you mean here … the homogeneity of the data used in a gridded dataset is the key issue. Several reanalyses have well documented discontinuities related to changes in data input streams.

We completely agree. Actually we meant the spatial homogeneity, but this should be captured by the previous part of the sentence. So we removed this part of the sentence that evidently created a misinterpretation.

8. Line 317: There is no season dedicated to the snow-cover onset period (Nov-Dec?), but there is one (March-May) for the spring season. Any reason for this? In my work documenting snow cover variability over Bulgaria (Brown and Petkova, 2007, Int. J. Climatol. 27, 1215–1229) and Quebec (Brown, 2010, Hydrol. Process., 24, 1929–1954) we found different trends in the fall and spring periods as well as different modes of atmospheric variability influencing snow cover variability in each season.

Our idea was to use the single months as an alternative to seasonal aggregates. However, in the conclusion we aggregated by season (cf. your comment 11.), so this was not very consistent.

Furthermore, we decided to include analyses of mean seasonal values of HS, maximum HS, and SCD in the manuscript.
However, we will place most of these in the appendix, because the paper is already quite long. The most important results were retained for the conclusion table.

9. Line 400: Can you please include the variable(s) the trend is computed for to remind the reader what the results refer to.

Good idea. We modified the sentence accordingly.

10. Lines 473-474: "In relative terms, the elevations of the stations used in this study oversample the elevations up to 1000 m, are similar from 1000 to 2000 m, significantly underrepresent 2000 to 3000 m, and do not cover elevations above 3000 m". This begs the question of why you did not attempt to transform the observations to an equal area grid (e.g. by kriging, or pseudo obs from modelling) to force the spatial coverage to be representative prior to the PC analysis?

This is a good point. But not a trivial issue. Because of the complex topography and strong elevational gradients in the Alps it is difficult to choose an appropriate resolution for the interpolation onto a grid. This would require a balanced number of stations both horizontally and vertically - a condition which is not met across the whole domain with our station set. While the transformation of the observations into an equal area grid would be interesting, we think it goes beyond the scope of the manuscript.

The main intent of this section is to give an overview of the confidence that can be expected from our assessments with respect to spatial coverage and elevation. We clarified it in the manuscript.

11. Conclusions: This study provided useful new insights into snow-climate regions and trends for the European Alps, but did not look at interannual variability in snow cover e.g. PC analysis of annual time series of snow cover duration and maximum accumulation. Is there a particular reason why you chose to ignore this? Documenting and understanding interannual variability is a key component of interpreting long-term changes (e.g. the signal-to-noise ratio of climate heating induced changes).

We followed your suggestion and included a section on interannual variability, as well as analyses on SCD and maximum HS; see also our replies to comments 6. and 8.

Reviewer 3

General comment. I was delighted to see this compilation and analysis of snow records from the whole span of the European alps. Previous country-based studies have used different methods that prevented aggregate conclusions, and the efforts the authors have undertaken to compile this comprehensive dataset represents an important breakthrough that
370	paves the way for a much improved understanding of the consequences of warming for snow in the European alps. Having assembled three datasets (with more similarities than those here) from different jurisdictions for some of my work, I can appreciate the magnitude of the task.

Thank you for this positive assessment which we highly appreciate. Please find below a detailed account of our changes to the manuscript as well as responses to your remaining comments.

375	Two referees have provided some technical corrections, to which I add the following.

Abstract - lines 49-51 are an attempt to represent much of the information in table 3 in a line of text, but the result is insufficiently specified and confusing. I suggest reducing the amount of detail and focusing on the key numerical message, and delivering it clearly. Perhaps one number for the DJF all-station average and one for the MAM all-station average. The next level of detail would be to list the average trends by elevation bands, but it's less confusing to put the elevation band
380	first: "for 0-1000m, -1.1cm/decade; for 1000-200m, …" Including the ranges is too much detail for an abstract, and places undue emphasis on outliers.

We adopted your suggestion to give all-station averages without ranges.

Regarding the elevational detail, we agree that we tried to condense too much information in too little space. We therefore remove numeric results from the abstract and now we only provide indicative remarks.

385	IPCC 2019 - follow the citation convention specified at the beginning of the report

Actually, we already took the citation as specified in the report. The IPCC 2019 citation refers to the SPM (summary for policymakers) part, and the citation has only been adapted to journal rules.

There almost seems to be a straight line through the loadings of PC2-5 (Fig 3) at about 47.5°N, straighter than the topography would suggest. It's suspiciously close to the Germany-Austria border. Can you convince me that it's not a data
390	artefact?

Yes, this impression can arise. And it is possibly also strengthened by our choice of the color scale. But our analysis indicates that this is not a data artefact. The issue is rather that the border between Germany and Austria is tightly linked to a strong topographic divide. It is also the case e.g. for France-Italy, where also PC3-5 can give such an impression. We added country borders in the topography map Fig 1(a), to make it easier to see that the change is linked to topography
395	rather than national borders.

Also, we created the PC figure splitted by country, in order to provide a clearer overview on the different data sources, and to better highlight the fact that the gradient is not a border effect (here only the subset for Switzerland, Germany and Austria for PC2):

PC2
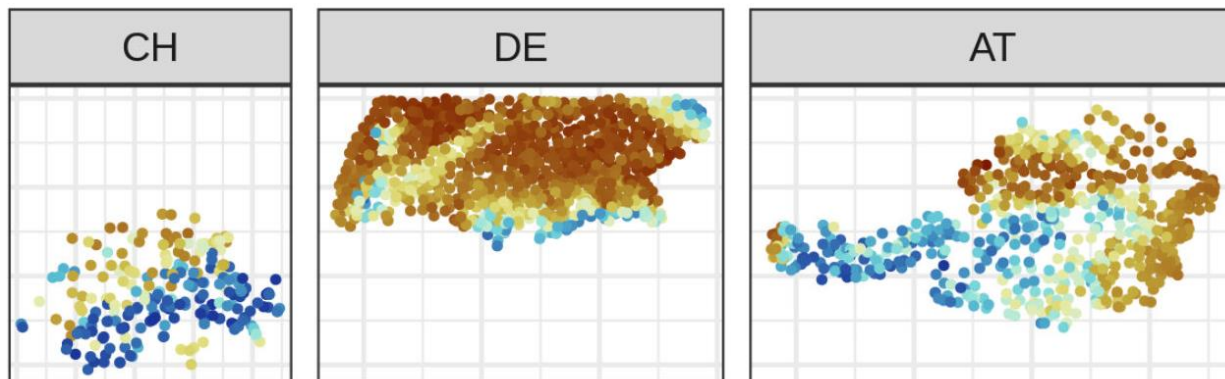
We shall add the full figure with all PCs and countries in the supplement, and discuss this issue in the manuscript as well.

Fig B1 is very important for the interpretation of the loadings; I strongly suggest moving it to the main paper

Done.

Line 401: state the p-value of significance

Done.

Section 3.3 - I see no real reason to shorten the record and present 30-year trends, except to calibrate the variability of shorter trends. I see another reviewer provided extensive comments on this.

We remove this section from the analysis (see also the detailed information on the response to Ross D. Brown review comments).