Summary and general comments

In this study, a statistical bias adjustment and downscaling method for snow cover fraction data is explored. Simulated snow cover fractions are obtained from the EURO-CORDEX regional climate model ensemble at a horizontal resolution of 0.11° – both for the present day climate and for the end of the 21st century (for RCP 2.6 and 8.5). The bias adjustment is performed with MODIS remote sensing data, which is spatially aggregated from a resolution of 250 m to a similar scale as the EURO-CORDEX data. For the bias adjustment, four different methods are initially applied – delta change, quantile mapping, quantile delta mapping (QDM) and multivariate QDM. Bias-adjusted snow cover fractions are then subsequently downscaled to the initial high resolution of 250 m. The entire procedure is, for a subdomain, compared to a rather conventional approach in which snow cover fraction is derived from a high-resolution snow model, which is forced by downscaled RCM output.

This study explores an interesting and novel approach – namely the application of bias adjustment for snow cover fraction. The authors evaluate the suitability of different bias adjustment methods for this challenging approach. The manuscript is predominantly well written – some sections, like the explanation of the downscaling method – are however difficult to follow and should be improved. At some places, a more extensive discussion of the results would be useful – for instance regarding projected relative changes in snow cover fractions – which are surprisingly largest at high elevations.

Thank you for the positive appreciation of our work and your extensive review, which can greatly improve a revised version of the manuscript.

Major comments

Remapping of high-resolution MODIS and CORDEX data
I'm a bit confused about the performed remapping. You first aggregate MODIS data to a low resolution grid and then remap CORDEX data to this grid with a nearest neighbour method, right? This seems to be unnecessarily complicated. Wouldn't it be easier to remap MODIS data (for instance with a conservative method) directly to the default rotated latitude/longitude CORDEX grid?

Yes, your proposition seems at first easier. However, it has been complicated by computational constraints. The support for the rotated CORDEX grid is very low in the software we used, namely R statistical software. At the start of the project (2-3 years ago), it was complicated if not impossible to deal with the rotated grid in R appropriately. Nowadays, support is a little better, but still not good. So, we decided to take the more standard LAEA grid of MODIS as backbone and reproject RCMs.

Downscaling method
The downscaling method is very interesting but difficult to follow in some places:

- A general question (just out of curiosity): downscaling can also be performed directly within the bias adjustment method. However, it seems to be less appropriate if observational data is available on a much higher spatial resolution. Did you nonetheless consider this option?

Yes, our initial idea was also to directly apply QM or QDM with observational data. In our case, however, we had to apply it to different types of variables: At 0.11° RCM scale we have a continuous but bounded snow cover fraction (0-100%) and at 250m observation we have a binary snow/land variable. So, we could not apply QM and QDM directly. On a side note: While the MODIS product we used is a binary one that has been developed at the institute, there exist also other options. Eg. the MODIS V6 product from NASA offers the NDSI (normalized difference snow index), from which snow cover fraction could be calculated. So with other observational data, the bias adjustment could be combined with downscaling directly. This would be an interesting alternative, but beyond the scope of this study...

- Line 229: Is the requirement for the function not rather "strictly increasing" to get a unique solution for SNCp50? Furthermore, I don't understand the subsequent part with the "longest non-strictly increasing subsequence"

Actually, non-strictly increasing suffices. SNCp50 occurs when the y=0.5 line is crossed (e.g. Figure 2). For

this, it is not important if we have a flat (non-increasing) portion. What destroy the uniqueness is when the y=0.5 is crossed twice or more times, which happens if the curve is decreasing at some point.
Regarding the second point, if the curve is decreasing at some point or even at multiple points, then there are multiple choices to select a monotonic subsequence. So we selected the longest possible of these, assuming it is the best estimate. We can provide an additional explanatory figure in a revised manuscript, if required.

- I found it particularly difficult to follow the "filling of missing SNCp50 values" section (lines 235 – 248). The comprehensibility might be improved with an additional sketch or figure. Moreover, I have some specific questions:
  - Could you explain why the Wasserstein distance is a suitable metric for this problem?
  - What do you mean by "no one-to-one correspondence"?
  - By nearest low-resolution pixels (line 240), you refer to the Wasserstein distance, not the horizontal distance, right?
  - Line 247: how do you determine these 100 pixels exactly from the theoretically 50 * 100 available pixels?
  - How did you proceed with the very small fraction (<0.001%) of pixels with still missing SNCp50 values?

That is a good idea. We can provide a graph/figure in a revised manuscript that gives an overview of the procedure, and we can also improve the writing in general. We can clarify all of the below points in a revised manuscript.
Regarding your specific points:

- The Wasserstein distance works for two sets of values that do not have to have a one-to-one correspondence between observations, such as, for example, probability distributions. The elevations of high-res grid cells can be considered such a "probability distribution". If we used other metrics, such as Euclidean distance, then we need to have one-to-one correspondence between the high-res elevations of two low-res grid cells. Such a correspondence could be e.g. by always comparing the topleft high-res grid cell of low-res grid cell A with the topleft high-res gridcell of low-res grid cell B, … , and the bottomright to the bottomright. But this would create not useful distances for the task, which is to compare the sub-grid elevation distributions.
- One-to-one correspondence: see previous comment.
- Exactly, at line 240 we refer to the Wasserstein distance.
- At line 247, we order the pixels by elevation difference.
- The rest of the <0.001% pixels were removed from the analysis.

- Line 249: Could the downscaling approach not simply be evaluated by reconstructing high-resolution MODIS snow cover from the spatially aggregated pixels?

That is true. We could also apply it directly to the upscaled observed MODIS data. We can explore this analysis in a revised version.

- It seems that you assume a seasonally stationary downscaling relation, i.e. there is no temporal dependency. Is this assumption valid? Or could the relation look slightly different for e.g. late autumn and spring?

Yes, we assume a seasonally stationary downscaling relation. The main reason is that the main components of change in snow cover fraction are snow deposition driven by snowfall and snow disappearing from snow melt. The strongest element in both snowfall and melt is elevation - at least at the scales used here in the 250m range. And this is what the downscaling procedure extracts.
Once we go up to higher resolutions, such as 10-20m, where preferential deposition of snow, terrain shading, and wind start to play strong roles - then the seasonal stationarity is questionable, since it will strongly depend on whether it's the start of the season or end of the season.
One idea to test the seasonal stationarity assumption is to derive two different SNCp50 values, one for the first half-year (e.g. Aug-Jan) and one for the second (Feb-Jul), and see whether they differ systematically.

Projected changes in snow cover fraction for different elevations

I'm surprised that the elevation gradients in Figure S12 are so distinctively different. Could this also be caused by the different RCM ensembles used for AMUNDSEN and the statistical approach (i.e. do different RCMs for instance show different elevation gradients in warming)? It is generally a bit counter-intuitive that the largest relative reduction in snow cover fraction is projected for the highest elevations (see Figure S10b and S12).

Yes, we too were surprised about the large differences in the elevation gradients in Figure S12. While beyond the scope of the study, it would be very interesting to understand where these differences come from. However, this would require a different experiment setup (same RCM-GCMs, with and without bias adjustment, …). Regarding your questions: Yes, different RCMs show different elevation gradients, see e.g. Fig S5 or S8. But while the model ensemble affects the spread, for the mean signal, the difference between the ensembles used for AMUNDSEN and the statistical approach should be little.
In Figures S10b and S12, we show the difference in percentage points, which is the absolute change in snow cover fraction, whose units are already percent. These are not true relative changes. In addition, the changes are for the annual snow cover fraction. While in winter largest changes are found for low-middle elevations, in spring, and likely summer too, the largest changes are for higher elevations - see e.g. Fig S5. In a revised manuscript, we could show a seasonal version of Fig S12, as well as a true relative change figure. These could then maybe explain better why the absolute changes are highest for higher elevation. We can also clarify better the difference between absolute changes in percentage points and relative changes in percent.

Minor comments

Content-related (text)
Line 15: With which bias adjustment method were these number computed?

QDM. We can report it in the abstract, too.

L17: "The comparison of the statistical" This evaluation confirms the robustness of the downscaling method I would therefore mention it before presenting results for the future climate.

Makes sense.

L19: The term "bias correction" is a bit outdated. Better use the term "bias adjustment" (see e.g. https://cordex.org/data-access/bias-adjusted-rcm-data/ and https://hypeweb.smhi.se/what-is-bias-adjustment/)

Thank you for pointing this out. We shall change the terminology throughout the manuscript.

L19: "plausible" might be a bit vague. Maybe better "more suitable" or "applicable".

"More suitable" sounds better.

L37: I find it difficult to follow this sentence. Maybe it's better to write "because climate change violates the assumption of stationarity" (if you mean that).

That's what we meant. We can rewrite it.

L53: "…and even snow depth" this part seems to be a bit contradictory to the previous sentence. Snow depth is reproduced relatively well but SWE not?

Yes and no. We could mention factors that might lead to these contradicting results, such as that in-situ snow depth observations are of generally higher quality than gridded SWE estimates, or that it could depend on the region.

L54: "gridded estimates of SWE" You refer to data derived from in-situ measurements (and not models), right? If so, I would explicitly state this. Furthermore, I guess you want to allude to the spatial representativeness of different data sets (in-situ vs. remote sensing) here. If so, you could discuss this in an additional sentence here…

The SWE grids we mention here are based on in-situ and remote sensing data, with some (empirical or physical) models behind. But we could also mention in more sentences the spatial issue.

L61: "The reference observations can be points or grids, are often limited in extent compared to RCMs, and feature, in case of grids, typically higher resolutions."

Thanks for the proposition. It sounds better.

L82: "The motivation is that…" this sentence is rather long and difficult to read – could you rephrase it? I furthermore do not understand the end "…while future change estimates should be consistent." Consistent with what?

Consistent across time. So, if we assume systematic biases, these should be consistent across time. We can rephrase and shorten the sentences.

L84: "While bias correction…" I don't understand this sentence entirely. What is the link between absolute values and bias adjustment?

Here we meant bias adjustment with downscaling. We can rephrase it.

L118: Did you apply a nearly or entirely cloud-free data set in the end?

It was a nearly cloud-free data set.

L119: What does "nominal" and "effective" resolution mean in this context?

Nominal is 250m, effective in the LAEA projection is 233m.

L127: "all variables" which variables do you mean?

While the coverage for temperature and precipitation is high, snow variables such as snow cover fraction are not available for all models.

L139: I was wondering – would it also be possible to account for the snow accumulation issue within the bias adjustment method?

We think not. If we assume a fully snow covered grid cell across the whole simulation period, this would likely render the change estimates for this grid cell inadequate. For SWE, the change would be too positive, for snow cover fraction, no change would occur. The bias adjustment might pull the hindcast closer to observations, but it cannot "correct" for this accumulation for the future.

L160: What do you mean by "single GCM-RCM biases"?

We meant that QM removes the biases for each GCM-RCM for the past period, and thus also the comparison for future estimates will be affected.

L171: I would be careful that the reader is not confusing calendar with hydrological years. I guess you conduct all your evaluations for hydrological years – right? Then I would state the definition of a hydrological year somewhere and mention, that all years refer to hydrological years – unless otherwise stated.

Actually, we performed calculations for the RCM data for calendar years. The bias correction has been performed month by month anyway and with distributions. Only for calculating snow cover duration from the incomplete MODIS data (Feb 2000 to Aug 2020), we used hydrological years. We can clarify this.

L175: I would remove ", usually 1970 – 2000"

Ok.

L189: “also to the maximum of 1.” this is oddly phrased. I guess values above 0.999 are set to 1.0, right?

Exactly.

L188: The trace condition was applied to the about of the Q(D)M algorithm - right?

Yes.

L195: “breaks the temporal consistency in the bias corrected SNC time series” what do you mean by that?

The random component is added in a distributional manner for each month and 30 years of data. So for climatological averages for that month, that should be fine. But at the daily scale, this random component can cause, e.g., that even during the melting period snow cover fraction can increase. And also the year-to-year variability might not be representative.

L260: What do you mean by “preferred”? Are trends only preserved for ratio variables?

No, for QDM trends are preserved also for the interval variables. What we meant here was that, the trend-preserving attribute is preferred for ratio variables in general.

L388: What do you mean by “analogue based”?

We meant climate analogues, which is a downscaling technique that looks for similar (analogue) observations in the past that might be in different locations. We can explain this further in a revised manuscript.

Typos, phrasing and stylistic comments

General: Both the terms grid cell and pixel are used. To avoid ambiguity, it would be good to use them consistently: You could for instance use grid cell exclusively for climate models and pixel for remote sensing data.

Good idea. We shall employ consistent naming in a revised manuscript. We can also include all suggestion below.

L22: remove “only”
L28: “consequences in water supplies” sound odd. Maybe rather “on”?
L31: maybe better “snow cover causes a significant atmospheric feedback due to its high albedo” **L50:** “such as for” sounds a bit odd. Maybe better: “for instance in”
L51: “but also from biases in RCM snow cover.” I would rephrase this to something like: “but also from biases caused by the relatively simple snow schemes of RCMs.”
L67: I would rearrange this sentence to: “Since QM has been show to modify trends in a few cases, quantile delta mapping (QDM) was developed, which represents a trend-preserving QM approach.” **L77:** “We restrict the study to snow cover fraction, which is, in contrast to snow depth and SWE, globally available on a high spatial resolution. The presented method has therefore a global potential for application.” **L82:** “meteorology” “output”

L88: I would replace “representation” by “pattern”. “Representation” can be ambiguous in this context (one could understand that snow cover representation in RCMs is improved).
L91: “is as” “is structure as”
L95: The study region (Fig. 1) encompasses the European Alps and spans from ~43 – 48.5° N and ~5 – 17° E, which roughly corresponds to the Greater Alpine Region (Auer et al., 2007).
L133: “although snow depth match very good” “although snow depths are well reproduced”
L172: “merged to” “merged with”
L270: “under- and overestimating” “simulated”
L305: change “lower ensemble” to “smaller ensemble”
L325: change “a more of a fully snow free and fully snow-covered grid cell over time” to “more often

fully snow free or snow-covered conditions over time”
L363: I would not write “became” if you compare two hypothetical projections
L383: Change to: “In contrast, the final imputation step of SNCp50 is based on a simple elevational dependence of snow cover, and could thus directly be estimated from a low-resolution RCM signal.”
L396/397: “in low/higher elevations” “at low/higher elevations”
L407: “meteorology” “output” or “meteorological output”
L470: What do you mean by “two months short snow cover duration”? A shortening by two months? **L474:** “lower horizonal spacing” “higher horizontal spacing”

Figures and Tables
Figure 3: Why is there no line in the upmost row in column “DC”? Furthermore, are results shown in this figure computed with bias adjustment and downscaling (or only the former)?

DC has no past observations of its own. They are actually the same from the RAW column. And results here are from bias adjustment only. We can explain this better in the caption.

Figure 6: Interestingly, the line for “bias correction” is closer to “AMUNDSEN” than “downscaling” in the rightmost panel. Can you explain this?

This is the case for RCP8.5. For RCP2.6 this only holds for > 2500m, and below downscaling is closer to AMUNDSEN. Unfortunately, too many factors differ that we cannot test given the methodological setup to answer this question.

Table 1: I’m only partially convinced about the point “+ Consistent climate change signal of surface meteorology with snow cover”. This is true for the raw RCM output, but no longer for the bias corrected product.

True. We shall remove this point from the table.

Figure S1: Replace “snow plausible” in caption with e.g. “snow accumulation”

Fine.

Figure S2: “surrounding” is a bit misleading in the caption – maybe better “aggregated”

Yes, aggregated would work. Or encompassing.

Figure S11: caption line 2: “and downscaling.”

Yes.

Figure S12: I guess the part in the squared brackets belongs to the figure caption?

Yes, it was copied from Fig. 6, but we can put it to the rest of the caption of Fig S12.

Table S5: I’m not sure if the reference “See also Figure S6” is correct/intended.

Thanks for spotting this typo. We meant Fig S7.