Response to Referee 3

We thank R3 for this detailed review, which will enable us to significantly improve our article. Enclosed please find a detailed explanation of the revisions we made based on R3's comments. For your convenience, comments are in bold and our response is in Arial italic. Revisions we made in the manuscript are presented in Arial italic with grey background.

General comments

The paper is generally interesting and could provide useful tool to adjust climate data on mountain regions. Especially the adjustment of meteorological variables which affect snow depth is highly welcomed as the accumulation and melting of snow is usually difficult to reproduce even on the areas with relatively constant altitude. Although the paper is promising I have some major comments which I think should be considered before publishing:

We thank the reviewer for this review, please see our specific responses to each point below.

1. This paper was hard to read and in some parts to understand as it uses difficult language and too long sentences.

We sent our article to a professional English translator who helped improve the langage.

2. Too many figures. Authors should reduce the amount to half at the actual manuscript and really think through what are the most important figures essential in supplementary material. Despite the authors' good intent the supplementary material with 207 figures is too much. Authors can not assume any reader to have time or willingness to go through those all. It is not good practice to refer to figures 1-207 (!) with every result authors show. 1-3 figures per result should be enough. Font size in figures is too small. It is not stated in every figure caption are the values hourly/daily/monthly/seasonal mean values. "Mean precipitation" does not tell much when the reader is not familiar with the study area and its climatic features.

This is a point that was shared by all three reviewers. We decided to remove figures concerning the Northern and Southern Alps, to keep only figures showing results for the Vercors massif as an example (with larger fonts and better quality) + the same figures for every massif in the French Alps in the Supplement. In the main article, we now have 15 figures instead of 28. Moreover, we decided to include a new synthetic table (Table 3) showing different features (mean values, biases, RMSE values and correlations) for variables of temperature, precipitation and snow depth for every massif in the French Alps + the Northern and Southern Alps, for the « RCM L. 1980-2010 » simulation configuration, at 1200 and 2100 m.

3. Authors state they have adjusted also wind speed, humidity, and short- and longwave radiation but do not show any results for these variables. It would have been interesting to see how large effect these variables actually have on snow depth and how much does the bias correction improve the results. This is especially interesting as authors have used hourly data where the variability can be larger than in monthly means.

Thank you for this remark. Concerning your first statement, we have focussed in this study on variables known as the most important in the study of the snowpack. We think it would be very interesting to look at other variables in a future paper. To see how much the bias correction improves the results would mean using outputs from uncorrected RCMs, which would not be on the same domain, making the direct comparison difficult. Moreover, driving impact models with uncorrected RCMs can lead to inconsistent results.

4. Although the quantile-quantile mapping is well known it is unclear how it is implemented in this study. Especially how is the extreme tail of distributions (>99.5%) handled? Add a short description and clarify the description of ADAMONT method as it is currently hard to follow.

Indeed, we thank R3 for this remark, which was also shared by the other reviewers. A better description was inserted in step 4 & 5 (l. 218-228) :

« 4. The quantiles (99 percentiles + 0.5 % and 99.5 % quantiles) of the RCM distribution and the SAFRAN distribution are then calculated at each centre point of each massif and for each elevation band, for each variable, each season (DJF, MAM, JJA, SON) and each of the four weather regimes.

5. Quantile mapping is then applied to the entire RCM dataset for the period 1980-2010, taking into account the season and the weather regime. For the values between quantiles, a linear interpolation is used. For RCM values greater than the 99.5 % quantile, a constant adjustment based on the value of this last quantile is applied. For precipitation, it can happen that for low quantiles, the probability of precipitation is lower in the RCM than in SAFRAN (i.e. several null values in the RCM, which can correspond to different positive values in SAFRAN). In this case, a random draw is performed amongst the SAFRAN values within the same quantile.»

5. Why are RCM's daily values disaggregated to hourly if all results are still presented as daily/monthly/seasonal mean? Authors should make it clear why the hourly data is important for this study.

Sub-daily data is necessary to drive impact models such as the SURFEX/ISBA-Crocus snow model presented in this study. However, you're right that we haven't emphasized this point enough in the current manuscript. The beginning of step 7 (l. 239-240) was changed to take into account this remark :

« The adjusted RCM dataset is then disaggregated from a daily integration period into an hourly time step, necessary for driving impact models such as the SURFEX/ISBA-Crocus model »

6. Why are the results shown as mean values for larges areas if the downscaling/adjustment is done separately for each massifs? This smooths especially the extreme values from the data and hides partly the true performance of the method.

Results are shown for the Vercors massif as an example, plus for each massif of the French Alps in the Supplement. The choice of showing results for the Northern and Southern Alps can be justified by the spatial scale of the effects of climate change, which is generally regional, and by the use we want to make of this method, i.e. to assess future conditions, whose stakes are generally expressed at the scale of the Northern and Southern Alps. However, as explained in a previous response (R3 General comment n°2), we have decided to show only results for the Vercors massif in the figures (+ for all massifs in the Supplement), and to limit the interpretation to the Northern and Southern Alps in a synthetic table (Table 3).

7. Be consistent with the names and definitions throughout the paper. Be spearing with acronyms especially if those are used only once.

This was taken into account.

Specific comments

Lines 67-69: As far as I know the quantile mapping is restricted to the range of observations unless there is added some method to handle the larger values than what was found from the learning period.

Yes, we thank R3 for this remark. We have added some explanation to take it into account (and more explanations about how we deal with the distribution tail later in the document) :

l. 67-71 :

« Moreover, the adjustment is not strictly restricted to the range of observed values in the reference period, which is the case for example for methods based on analog weather patterns (e.g., Déqué, 2007; Themeßl et al., 2011; Rousselot et al., 2012; Dayon et al., 2015), provided that values based on the lowermost and uppermost quantiles are handled appropriately (Gobiet et al., 2015). »

Lines 109-112: State clearly that only past climate is studied in this study. I thought also future period was considered here.

R3 is right. The method is intended to work also on projections of future climate, but for the evaluation we focussed on recent climate. We adjusted the following sentence (l. 120-122) to state this :

« In order to evaluate the performance of the ADAMONT method, here we apply this method to the ALADIN-Climate v5 RCM (Colin et al., 2010) forced by the ERA-Interim reanalysis (Dee et al., 2011) over the period 1980-2010. ».

Lines 119-120: what about section 4.

Indeed, we forgot to indicate the Discussion. This was added in the new version of the manuscript.

Line 165: centroid = center point of some grid point or massif area?

Yes, it is the center point of each SAFRAN massif. We have changed the nomenclature in the whole manuscript.

Line 195: What is the range of elevation factor N (0-1, 100-1000 etc.)? How it depends on the altitude?

This question is related to a remark by R1. We tested values of N of 50 and 100, and only showed results for N=50, as explained later in the document. This was empirical : using N=50 yielded satisfying neighbouring grid points, while N=100 yielded neighbours that were sometimes too far from the SAFRAN centroids (or center points). N does not depend on the altitude, it is simply a factor we use to give more weight to the proximity between grid points and SAFRAN center points in the vertical direction than in the horizontal ones.

We inserted the values of N we used in equation (1), l. 196-198:

« (…) and N is referred to as the elevation factor. Values of 50 and 100 were tested, but only results with a value of 50 (N50) will be shown in this study. »

Lines 201-202 and 214-216: Why is the hourly SAFRAN data first integrated to daily, then used in the quantile mapping function with RCM daily data and then the adjusted RCM daily data is disaggregated to hourly using the same hourly SAFRAN data? Why isn't the RCM data disaggregated to hourly before the ADAMONT adjustment?

This was done to keep the consistency of the daily cycle of each variable. Using the analogue technique presented in step 6 to disaggregate the RCM to hourly before the ADAMONT adjustment would give inconsistent results with more frequent discontinuities, because SAFRAN and the original RCMs are too different from each other. Moreover, SAFRAN and other reanalyses projects (UERRA, EURO4M, see for example Soci et al., 2016) are generally more relevant daily, because some of the observations assimilated in the system are only available at a daily time step (for example, precipitation).

Lines 201-202: These integration methods are not clear to me. Have those been shortly explained somewhere? Table 1, method column.

R3 is right, those methods were not really explained in the manuscript. In fact, the same types of methods are used in step 3 for the integration from hourly to daily and in step 7 for the disaggregation from daily to hourly. R1 asked for a clarification of step 7 regarding those methods. But we should also clarify step 3 accordingly. We added the following short explanation about the integration methods in step 3 (l. 212-217):

« The SAFRAN data are integrated from hourly to daily time resolution to match the data content of the available RCM output. The integration method depends on the variable considered (see Table 1) : for temperature, the daily (6 am to 6 am the next day) minimum and maximum values are selected, for wind speed and humidity, the last value of each day (at 6 am) is selected (in order to be comparable to an instantaneous value), and for precipitation and radiation, the daily mean is used. »

Lines 206-207: Here time period for RCM is 1980-2010 but in figures is used 1979-2010. Why? Describe shortly the quantile mapping method used. There are different variations depending on the treatment of extreme values.

Concerning the period (1979 or 1980 – 2010), we thank R3 for this remark. We forgot to correct the period in the figures and elsewhere in the text, it should be 1980-2010. Moreover, the two half-periods should be 1980-1995 and 1995-2010. This was corrected throughout the manuscript.

Concerning quantile mapping, please see our response to general comment n°4 above.

Lines 208-214: I did not understand this step. Is this step 6 supposed to clarify the step 4 or to precede the step 7? Is this step done daily or monthly? And how does it differ from the step 4 where seasonal percentiles were calculated?

This step precedes step 7, in which SAFRAN hourly series of daily analogues selected for each RCM day in this step are used. It is thus done daily.

Indeed, this step was not clear enough. We improved its explanation in the new version of the manuscript (l.229-238) :

« 6. For each day in the RCM dataset, an analogous date is chosen in the SAFRAN dataset, matching the following criteria: the month and the regime must be the same as in the RCM dataset, mean precipitation over the Alps must be consistent between datasets to ensure intermediate-scale (accross the French Alps) climatological consistency (i.e. if precipitation in the adjusted dataset is less than a threshold of 1 kg m −2 day −1 , precipitation in the SAFRAN analogue must also be less than this threshold), and whenever possible, consecutive time slices are chosen in the SAFRAN dataset in order to avoid artificial jumps in the final data linked to the choice of analogues. For each RCM date, a random draw amongst all available SAFRAN dates is performed, then the dates are browsed through until one meets all the requirements. This analogous day is then used in step 7 for all variables. »

253-254: In Olsson et al. (2015) they found that the separation of temperature to dry and wet days produced unrealistic results compared to observations and they used unseparated temperature data for the final results. Was any comparison made with and without separation of precipitation to rain and snow?

Yes, in the results (Figs. 5-6), we use a specific configuration (no corr) where we look at total precipitation (rain + snow) and snow depth without performing the last quantile correction on rain and snow separately. In fact, the separation of precipitation into rainfall and snowfall is an information that needs to be given as input to the SURFEX/ISBA-Crocus snow model, so we decided to produce it anyhow.

Lines 269-271: How much does this grouping decrease and smooth the extreme values?

We generally look at mean values in the results, not at extreme values. Thus this is out of the scope of this article.

Lines 277-278 (throughout the paper): It is not a good practice to ask the reader to go through 207 figures to get some clue what would the conclusion of the results be.

The reader is not asked to read all 207 figures, instead he can focus, in case he/she is interested, on the specific massifs he needs. But you're right that we shouldn't put a reference to « Figs S1-S207 » . We removed all these references to keep only « Supplementary Information »

Lines 184-301: Why slightly different periods 1979-2010 and 1980-2010 are used in results?

Please see our response above (R3 specific remark Lines 206-207).

Lines 293-300: Does this mean the relative proportions of wet and dry days are calculated from the whole period separately for RCM and reanalysis and then used to calculate the specific scores or were these calculated so that it had to be dry or wet in both RCM and reanalysis at the same day(hour)? In RCMs the relative proportions should be similar to observations/reanalysis after adjustment but the same weather will probably not occur in the RCM and reanalysis at the same day.

We calculated these scores so that it had to be dry or wet in both RCM and SAFRAN at the same hour. However, the ALADIN RCM in the context of this study is driven by the ERA-Interim reanalysis. We are not in the case where a RCM is driven by a GCM, in which case your remark about the same weather not occuring at the same day applies.

Lines 313-324: Quite long sentence.

Yes indeed. We replaced this long sentence by a list of 4 points (l. 360-374).

Lines: 345-346: Why isn't the table 2 referred already in section 2.5?

Section 2.6 (and thus Table 2) is now referred to in the following sentence (l. 324-325) in Section 2.4 :

The two RCM grid points neighbour selection techniques and the three different learning periods (1980-1995, 1995-2010 and 1980-2010, see Sect. 2.7) were tested.

Line 349: evidenced=evidences?

Yes, this was corrected.

Line 354 (throughout the paper): Please be consistent with names, definitions and acronyms. Here "our method" is ADAMONT method?

Yes, this was corrected.

Lines 358-360: Why is the average precipitation lower in the longest time period compared to the shorter time periods?

This is explained in the Discussion. Please see our response below.

Line 367: 9 figures with sub-figures to display the results for RMSE is too much. Please reduce.

Please refer to our previous response on this point (R3 General comment n°2).

Line 371: Why? Is the variability of temperature lower in autumn compared to other seasons?

We have no explanation for this statement.

Lines 373-375: Why is that? Is there less data or too large distances between altitudes?

This is explained in the Discussion (Section 4.2 Impact of the spatial selection technique)

Lines 381-383: If the figures 7-12 also includes the uncorrected values then it should be stated in their legend.

This is the reason why we produced Table 2, so that we wouldn't have to explain the different configurations in each figure legend every time.

Lines 385-386: I think bias of 150mm/month sounds quite large. Could you give these over/underestimations as percentage values?

In Table 3, we have now included the mean annual values of temperature and precipitation, and the winter mean value of snow depth, so that biases can be compared to those mean values.

Line 392: Also this bias sounds quite large. See previous comment.

Please see our response to the previous comment.

Lines 401-402: Why the N50 degrades results at high altitudes?

This is explained in the Discussion (Section 4.2 Impact of the spatial selection technique)

Line 409: What does this integration time mean?

Integration time was meant as an integration window. This was corrected troughout the manuscript.

Line 425 (throughout the paper): Is there difference between SAFRAN and SAFRAN/Crocus?

SAFRAN is the meteorological reanalysis. SAFRAN/Crocus refers to the Crocus snow model driven by SAFRAN (thus every time snow depth is mentioned). But R3 is right that in this particular sentence, only SAFRAN should be used because we only mention temperature. This was corrected.

Line 277: adjusted RCM = adjusted with ADAMONT method?

Yes, this was corrected throughout the paper.

Lines 480-482: This is not surprising as the quantile mapping should adjust the learning period values close to observed values! There should be stated how close the ADAMONT methods gets the observational data on the learning period and why there will be greater differences on other periods.

Thank you for this remark, indeed this is the purpose of quantile mapping. We included more explanation in Sect. 3.2, where this feature is first noted (l. 527-530):

« Concerning the choice of learning period, the ADAMONT method performs better during a period which corresponds to its learning period, because quantile mapping adjusts the learning period values close to SAFRAN values, while the adjustment for other periods is based solely on the mapping function determined for the learning period. »

In Sect. 3.3 (l. 540-541), we included :

« Some significant differences appear when using different learning periods, as already noted in Sect. 3.2,(...) »

Line 489: What is DSCLIM?

We forgot to introduce it. It's an analog resampling based transfer function algorithm (Pagé et al., 2009). This was included in the text (l. 551-552).

Line 490 (and forward): What is figure 10.1?

Fig. 10.1 in Lafaysse, 2011 (therein).

Lines 525-528: These acronyms have been already defined in section 2.5.

Yes, R3 is right. We removed the explanation of acronyms.

Lines 546-550: Why are the precipitation underestimated with the longest time period compared to the other periods?

The fact that it is underestimated is not meaningful, it could have been overestimated.

Lines 550-554: How was the extreme tail of distribution handled?

Please refer to our answer above concerning quantile mapping (R3 General comment n°4) .

Lines 555-558: Again, the bias correction methods should perform like this and the result is not surprising. How much did these periods differ from each other?

We included the following (l. 620-622) :

« (…) revealed some significant differences when using different learning periods, linked to the use of quantile mapping, with the ADAMONT-adjusted RCM simulation (...) »

The difference between the two sub-periods is explained in Sect. 2.6 (l. 392-394) : « Different learning periods were tested to evaluate their impact: 1980-1995 and 1995-2010, which correspond to periods with contrasting meteorological (and snow) conditions linked to regime shifts (Reid et al., 2015) »

Lines 590-593: ultimate correction = bias correction of rain and snow separately? Please be consistent.

Yes, this was corrected.

Lines 613-614: No need to explain the acronyms again and again.

This was removed.

Lines 653-667: How does ADAMONT method treat the lowlands where there are no massifs? Gridwise?

In its current configuration (using the SAFRAN reanalysis over French massifs), it does not treat them, it is only run over mountain regions where the reanalysis is available. Note however that a France-wide implementation of SAFRAN exists, which uses different zoning than the massif approach in the mountain regions (Vidal et al., 2010).

Table 2: What is the "period considered"? Meaning of RCM APPR?

The period considered in the figures (1980-2010, 1980-1995 or 1995-2010). This was clarified.

RCM APPR is in fact RCM L. (L. stands for learning). We changed the nomenclature and forgot to correct it here. This was corrected.

Figure 18: It is hard to compare the figures as they have different scaling.

This Figure now corresponds to Figure 8. Indeed, the x-axis scalings are different between 1200m and 2100m subplots, but the point of this Figure is not really to compare the PDF distributions between two elevations, but rather to ensure that for a given elevation, PDFs of ADAMONT-adjusted simulations are close to the SAFRAN one, which is the case.

Figure 19: The scaling of these figures could be reduced as there is too much white background.

Yes, but we used the same scale for the supplementary figures. Some of them need such a scale.