Response to Referee 2

We thank R2 for this helpful review. Enclosed please find a detailed explanation of the revisions we made based on R2's comments. For convenience, comments are in bold and our responses are in italic. Revisions made in the manuscript are presented in italic with grey background..

This paper analyses the mean state and variability in historical and future snow conditions at a mid-elevation location in the French Alps. Analysis is based on output from a physical snowpack model (Crocus) driven by a regional historical reanalysis (SAFRAN) and a suite of historical and future RCM simulations. In situ observations from a nearby location are also used for comparison of the historical conditions.
The methodology used in the paper is novel and represents an improvement on previous studies. Results are generally well described compared to existing literature. However I believe there are several changes required and details to clarify before I can recommend final publication.

We thank the reviewer for this overall positive appreciation of our work and hope that our revisions and replies will address his/her concerns.

Specific Comments:

Table 1 obscures the fact that there are only 5 distinct GCMs simulations that sample natural variability. I think this should be pointed out explicitly.

We now represent Table 1 as a matrix of the different RCM/GCM combinations. (p.6)

Figure 2a-c: In the supplementary material for the RCP8.5 scenario there is a distinct change in the stddev as average snow depth becomes small. It's therefore unfair therefore to suggest that the stddev is stationary based on the RCP4.5 scenario. I think it would be fairer to show the RCP8.5 scenario in the paper and place RCP2.6 and 4.5 in the SI.

We agree that RCP 4.5 and RCP 8.5 provide different end-of-century responses of the snowpack for virtually all indicators. It is also true that in the case of RCP 8.5, the snow reduction is sufficiently pronounced that the standard deviation can no longer be considered stationary. However, we believe that it would be unfair to choose RCP 4.5 or RCP 8.5 and focus on either one or the other in the presentation of results and analysis. We intended in the original version of the manuscript to show RCP 4.5 in the main body of the article and provide RCP 8.5 results in the supplement, for the sole sake of saving space. Based on the reviewer comment, we suggest that both RCP 4.5 and RCP 8.5 should be displayed in the main body of the revised article, so that this does not give the reader any impression that we favor one of these two RCPs.

The fact that stddev declines through time for RCP 8.5 is now indicated : « Figure 2e displays values on the order of 0.08 to 0.11 m with decadal variability but no temporal trend from 1950 to 2100. Figure 3e, on the other hand, shows a decline of standard deviation with time, as SD becomes smaller. ». (p. 11 L. 24-26)

Figure 2d: The large relative contribution to combined model uncertainty arising from snowpack model multiphysics compared to RCM/GCM inter-model variability seems difficult to reconcile with plots 2a-c (I realize there is interannual variability in Figs 2a-c, while much of this signal is removed in the multi-annual average). Is it possible to present

15 year running means for the 13 RCM/GCM tracks shown in Fig 2c or at least for a representative subset of the 13 pairs along with Fig d (either in this same figure, or in a separate figure similar to how you subsequently separate quantile plots based on annual frequency data and multi-annually averaged data)? This may allow the reader to get a better sense of the relative spread in the ESCROC ensembles compared to the RCM/GCM inter-model variability.

We agree that the relative contribution of uncertainty displayed by Fig. 2d (2e in the revised manuscript) could not be directly associated with the spread of Fig. 2c because the 15-year running mean removes the high interannual variability. Therefore, we added an intermediate panel (Fig. 2d) which is simply the 15-year running mean of Fig 2c. We think that this helps the reader understand all the post-processing and the links between the different subplots.

The description of this Figure was modified accordingly in section 3.1 :
« Figures 2d and 3d present the 13×35 15-year running average values spanning all simulation members of Figures 2c and 3c respectively. This corresponds to the second statistical post-processing described in Section 2.5.2 which removes the interannual variability and allows an easier quantification of each source of uncertainty.
Figures 2e and 3e aim at apportioning the uncertainty in the time series of Figures 2d and 3d respectively, between the uncertainty arising from GCM/RCM inter-model variability (including model uncertainty and internal variability of climate at different time scales) and the uncertainty arising from the multiphysics snowpack model. For that purpose, the standard deviations of the 455 values of Figures 2d and 3d were computed for each 15-year window, and correspond to the total standard deviations of the SD. This is shown in black solid line in Figures 2e and 3e. (...) ».
(p. 11 L. 17-24)

Further to this point — in the text there are several times that you refer to the relative fraction of uncertainty contributed by snowpack model multiphysics as ~20%, yet this graph shows that is can be as high as 80%. Please justify the use of 20%. Is there a rationale for why the uncertainty due to snowpack model multiphysics is higher in the historical period, even though the combined model uncertainty remains fairly constant?

R2 is right that the value of 20 % uncertainty is only valid for future time periods. The answer concerning the fact that uncertainty due to snowpack model multiphysics is higher in the historical period is partly given p. 12 L. 1-3 : « (...) the historical period is affected by the varying number of available GCM/RCM before 1980 and by a potentially artificial reduction of spread over the 1980-2011 calibration period of the ADAMONT statistical adjustment method ».

Just after this sentence, we added this implication in the manuscript:
« This could partly explain why the uncertainty of GCM/RCM appears lower than the multiphysics uncertainty on the historical period, in combination with the deeper snowpacks in the historical period. ». (p. 12 L. 3-4)

Indeed the uncertainty also declines with time linked to increased snow scarcity, as already indicated (p. 11 L. 32-35): « The ESCROC component shows values ranging from 0.02 m to 0.07 m, exhibiting rather smooth fluctuations from 1950 to 2100 and a general decreasing trend, along with the general decreasing trend of SD over the considered time period (see below) ».

Figure 3 and 4: The quantiles from different RCPs overlap too much in these figures to discern one set of shading from the others. I suggest showing RCP8.5 only in these figures. The results from the other RCP scenarios are provided in tabular form in the main document which I think is sufficient (along with the plots of other RCPs in the SI).

We do not agree to show only one RCP in the main text, and prefer to show all of them together, with individual RCP plots in the supplement. There is no rationale for choosing either one the existing RCPs.

Figure 3: Why do the SOD and SMOD for RCP8.5 begin to encompass summer months? Is this an error with the calculation?

We thank R2 for this remark. Indeed, there was an error in the calculations of SOD and SMOD for years in which SD was never greater than 5 cm. The algorithm was corrected with implications for Figs. 4-5 and S3 & S6 , as well Tables 2 and 3.

P8.L15-21: Please clarify why you use a distance of +/- 1.37*sigma from the mean for the 17th and 83rd percentiles? Shouldn't the 17th and 83rd percentiles be 0.95 sigma away from the mean such that CV=1.9*sigma/mu?

We thank the reviewer for spotting this error. Indeed, the Q17 and Q83 percentiles correspond to ± 0.954 sigma distance to the median, we used an erroneous value in the original submission. All tables, graphics and text using Q17 and Q83 percentiles values from multi-annual average values have been updated accordingly. We are particularly grateful to the reviewer to have identified this error, which may have caused propagation of erroneous indications of variability and spread from our projections had it not been corrected during the review process.

P25.L28-29: Please rephrase in order to account for the relative component attributed to snowpack modeling errors in both future and historical periods.

This was rephrased : « (…), under the conditions of the Northern French Alps and after the middle of the 21st century, the uncertainty component attributed to the snowpack modeling errors alone is on the order of 20%, (...) ». (p. 25 L. 33 to p. 26 L. 1)

P27.L10: This statement depends on the RCP scenario. It is not the case for RCP8.5.

This is now clarified: «(…) and even increases in relative terms (until the middle of the century for all RCPs, and towards the end of the century for all RCPs except RCP8.5), (...) ». (p. 27 L. 13-14)

P28.L15-17: Could there also be a change in the mean density of snowfall occurring at the location?

The density of snowfall depends on temperature and wind speed during the snowfall. While this is an interesting hypothesis to test, we consider this to be beyond the scope of this article and to be addressed in a future study.

P29.L8-10: Please reword or justify the claim that snowpack modeling uncertainty is typically 20% when Figure 2d shows it can be up to 80%. I agree that it may have a smaller impact on trends.

The value of 20 % is valid only in the future. This sentence was rephrased : « Uncertainty arising from physical modeling of snow after the middle of the century can account to 20% typically of the simulation results ». (p. 29 L. 25-26)

P29.L11-17: The ADAMONT method was evaluated in a previous paper. Please be clear as to which aspects of these conclusions were accomplished in this paper.

The ADAMONT method was described and evaluated only on one RCM driven by a reanalysis in Verfaillie et al., 2017. Here we apply it to the EURO-CORDEX RCM/GCM ensemble spanning the historical period and future projections for the 21st century. This is the first published use of the method, so that it is the first evidence of the capability of the method to be used to adjust a large number of regional climate model results and provide consistent meteorological forcing data for the land surface model Crocus.

Further to this point, while you argue that your methodology is an improvement to previous studies that use delta change methods, your assessment of the results says that they, in

fact, agree with these previous studies. Under what circumstances might you expect to see differences? Is it possible to provide a direct comparison between your methodology and a delta change method for this location or to highlight statistics that would differ between the two methods?

We agree with the reviewer that our results are consistent with results obtained using delta-change methods, in French mountain regions as well as in Switzerland, as quoted in the manuscript (e.g. Castebrunet et al., Schmucki et al.). However, this consistency is only demonstrated for multi-annual multi-model trends on snow depth or snow water equivalent mean values. We strongly believe that our model chain, using the RCM chronology, will capture more appropriately potential changes in timing of precipitation, and as such could only be compared to raw output of RCM (e.g. Steger et al.). Differences would be expected under a situation where the chronology of precipitation would differ significantly in the future, because the delta-change approach would only modify the air temperature and rain/snow partitioning, but not the timing of the events. These changes in the multivariate chronology of meteorological events in the Alpine region have not been investigated in details until now to the best of our knowledge although their stationnarity is a requirement for the validity of the delta-change method. Furthermore, although our results do not exhibit significant changes in the interannual variability of the snow indicators, this is a result of our projections whereas it is only an assumption in the delta-change method.
While this could be interesting to demonstrate whether results obtained using delta-change approaches could still be employed for impact studies, we do not feel the need to perform such comparisons ourselves given that there are no arguments supporting that our approach could provide less appropriate results than a delta-change approach. We have no ressource to implement a delta-change method for the purpose of such a comparison, now that we have developed and implemented the full model chain described in this manuscript. We are, however, fully eager to communicate our data to other research groups interested in performing such a comparison in the future.

We added a paragraph at the end of Section 4.3 to explain this in more details :
« Many of the results discussed above indicate a strong consistency between our results and results obtained using delta-change methods, in French mountain regions as well as in Switzerland (e.g., Castebrunet et al., 2014; Schmucki et al., 2014). This consistency is shown for multi-annual multi-model trends on snow depth or snow water equivalent mean values, but cannot be assessed regarding the interannual variability because this is generally not addressed in these studies. The model chain implemented here, explicitly making use of the intra-seasonal and inter-seasonal RCM chronology, inherently captures more appropriately potential changes in timing of precipitation. Differences between the current study and studies based on delta-change approaches would be expected under a situation where the chronology of precipitation would differ significantly in the future, because the delta-change approach would only modify the air temperature and rain/snow partitioning, but not the timing of the events. These changes in the multivariate chronology of meteorological events in the Alpine region have not been investigated in details until now to the best of our knowledge, although their stationarity is a requirement for the validity of the delta-change method. Furthermore, although our results do not exhibit significant changes in the interannual variability of the snow indicators, this is a result of our projections whereas it is only an assumption when applying a delta-change method. More in-depth comparisons between outputs of delta-change approaches and direct adjustments to RCM output could be carried out in the future, but are beyond the scope of this article. ». (p. 27 L. 19-32)

Technical corrections/Suggestions:

I find the use of the phrase "annual-scale" a bit unnatural. I suggest using "annual indicators" as you occasionally do (P8.L15) throughout the paper.

This was corrected. (p. 1 L. 2)

Similarly, there are places in the paper where you might consider replacing the word "variation" with "change", "response", "difference" or "variability", but I'm having trouble articulating a clear rule to follow in this. Variation is frequently reserved to specify a very small adjustment.

OK. We corrected this throughout the manuscript.

Title: "Multi-component ensembles. . .."

The title was changed accordingly.

P1.L2: "This article investigates the climatic response of a series of indicators for characterizing annual snow conditions and corresponding meteorological drivers at 1500 m altitude in the Chartreuse mountain range in the Northern French Alps. "

OK. This sentence was modified as suggested by R2. (p. 1 L. 2-3)

P2.l30: "because they are newer. . .."

Done. (p. 2 L. 29)

P3.L2-4. Please rephrase these two sentences to make the typical delta-change approach clearer.

We rephrased : « (…) a pre-determined difference (delta) of temperature and/or precipitation values to an observation record, based on changes computed using climate models (either global or regional). This cannot capture combined changes in temperature, precipitation and other meteorological factors, in terms of magnitude of the fluctuations and their seasonal-scale and interannual variability. ». (p. 3 L. 1-3).

P8.L15: "Moments of multi-year averages: A running average of annual indicator values is computed (typically with a 15 year sample window), for a given RCP and for each GCM/RCM pair."

We changed the sentence to: « Moments of multi-year averages: A running average of annual indicator values was computed using the 15 year sample window, for a given RCP and for each GCM/RCM pair. ». (p. 7 L. 23-24)

P9.L1 "for 15-year windows around each future time period t and each RCP r"

We changed this sentence to : « for 15-year windows around each future time period t for the RCP r ». (p. 9 L. 12-13)

P9.L3: "i.e." in place of "e.g."

Done. (p. 9 L. 15)

P10.L19: "It highlights the significant interannual variability in observed, reanalyzed and climate model datasets."

Done. (p. 10 L. 27)

P11.L37: "which highlights the need for appropriate data synthesis methods". Please elaborate.

We added the following sentence : « Indeed, it is not possible to draw conclusions or make decisions on the sole basis of such a raw ensemble of individual scenarios. ». (p. 11 L. 15-16)

P21.L10: to widen

This was corrected. (p. 21 L. 4)

P23.L27: "By definition no performance metrics pertaining to annual variations can be computed between the adjusted climate output and either observations or reanalysis data, because the two are not designed to exhibit synchronous variations."

The sentence was corrected : « By definition no performance metrics pertaining to annual fluctuations can be computed between the adjusted climate output and either observations or reanalysis data, because the two are not designed to exhibit synchronous fluctuations. » (p. 25 L. 1-2)

P25.L4: "independent from the time period for calibration of the ADAMONT adjustment method (1980-2011). . .."

OK, this was corrected. (p. 25 L. 10-11)

P25.L9-10: "or applying the final quantile mapping separately to rain and snow precipitation in order to mitigate detrimental interactions between temperature and precipitation (Verfaillie et al., 2017). . .."

This was corrected. (p. 25 L. 16-17)

P25.L30: "Because the number of GCM/RCM model pairs was different for RCP2.6 (4) and RCP4.5 and RCP8.5 (13), we compared the statistics for indicators during the historical period based on the 4 RCP2.6 pairs alone, as well as the full ensemble of 13 GCM/RCM pairs."

Done. (p. 26 L. 3-5)

P26.L11: "similar statistics are found for these 4 model pairs as for the full ensemble of thirteen."

Done. (p. 26 L. 15-16)

P26.L24: I'm not sure what you mean by "snow-dry" seasons. Seasons without snow on ground or without snowfall occurring at this location at all?

We meant seasons without snow on the ground. This is now better explained: « (…) and more frequent seasons with barely any snow on the ground ». (p. 26 L. 28)

P26:L26-28: "The decreasing SD trend is also combined with a decreasing SWE trend (~ -6 kg m−2 per decade for RCP2.6, -18 kg m−2 per decade for RCP4.5 and -35 kg m−2 per decade for RCP8.5 over the period 2030-2090, Table 4) and a decreasing trend in duration of STED5 (as in Marty et al. (2017a)), STED50 and STED100 (Table5)."

We have replaced the sentence by :
« The decreasing SD trend is also combined with a decreasing SWE trend (~ -6 kg m−2 per decade for RCP2.6, -18 kg m−2 per decade for RCP4.5 and -35 kg m−2 per decade for RCP8.5 over the period 2030-2090, Table 3) and a decreasing trend of STED5 (as in Marty et al. (2017a)), STED50 and STED100 (Table S2). ». (p. 26 L. 30-33)

P27.L33: "This is all the more relevant in that none of the GCMs used for this study. . .."

Done. (p. 28 L. 18-19)

P28.L6-8: ". . .in contrast to previous studies (Durand et al., 2009a; Pepin et al., 2015). This result may stem in part from the fact that although elevation dependent warming is generally maximal in the fall and springtime, our target period covers mostly wintertime. Alternatively, this low enhancement factor could be due. . ..."

This was corrected (p. 28 L. 26-28).

P28.L26-28: "The multi-component ensemble framework makes it possible to account for the various sources of uncertainty and variability that affect future climate projections, some of which are neglected in both previous and ongoing climate change impact studies."

Done.(p. 29 L. 12-14)

P28.L28-32: Split into more than 1 sentence.

The sentence was split in two: « The multi-ensemble framework developed here draws on several RCPs (RCP 2.6, RCP 4.5 and RCP8.5), feeding several GCM model runs from the CMIP5 intercomparison exercise, which themselves feed various RCP model runs from the EURO-CORDEX downscaling exercise. Those are adjusted using the refined quantile mapping method ADAMONT against the meteorological reanalysis SAFRAN, making it possible to drive a multi-physical version of the energy balance multi-layer snowpack model Crocus. » (p. 29 L. 14-18)

P28.L32-36: "The method defines a series of annual snow and meteorological indicators that represent various aspects of winter snow conditions. . .."

We changed the sentence to :
« The method defines a series of annual snow and meteorological indicators that represent various aspects of the winter season (...) ». (p. 29 L. 18-19)

P29.L21: "exhibit similar statistics at the interannual and multi-annual scale as the full 13-member ensemble, ..."

OK. (p. 30 L. 5-6)

P29.L26: "but maintained interannual variability. . .. . . " Rephrase.

We changed the word « maintained » to « sustained ». (p. 30 L. 11)

P29.L28-29: "As assessed in this study, for this location, interannual variability is larger than inter-model spread for a given RCP scenario."

This was corrected. (p. 30 L. 13-14)

P29.L32-33: "the latter leading to frequent occurrence of ephemeral or nearly snow-free conditions at the end of the century."

OK. (p. 30 L. 17-18)

P29.L35: "For example, the change in mean snow depth"

OK. (p. 30 L. 20)

P30.L4: "this value changes very rapidly": I dislike the wording that it changes "rapidly" since the changes on Figure 5 are quite linear (except for STED100). Please rephrase. I suggest something along the lines of "the magnitude consistently increases along with global mean temperature reaching reductions of 80% beyond 4 °C of global warming.

OK. We have chosen the formulation suggested by R2. (p. 30 L. 22-23)

P30.L8-10: "These locations may be investigated in the future, based on the methodological framework introduced here and the data available in the SAFRAN reanalysis for the French Alps and Pyrenees (Durand et al., 2009b, a; Maris et al., 2009)."

OK. (p. 30 L. 27-29)

Figure 2 Caption: "c) Ensemble of Crocus model configurations driven by the 13 RCP4.5 GCM/RCM pairs; each GCM/RCM pair is displayed with a different color."

OK. The Figure caption was changed accordingly. (p.13)

Figure 4 Caption: "Ensemble spread in 15-year running mean (μ ± σ') of all GCM/RCM pairs for each scenario (HIST, RCP2.6, RCP4.5 and RCP8.5), along with 15-year running means of observations (1960-2016) and SAFRAN-Crocus runs (1958-2016) at CDP, for: . . .."

OK. This is now Figure 5. The Figure caption was changed accordingly. (p. 18)

**Figure 5 Caption: "Response of local meteorological and snow indicators to global warming level. Indicator response is computed as the difference of multi-annual means between end of century (EOC, 2071-2100), middle of century (MOC, 2041-2070), or beginning of century (BOC, 2011-2040) and the reference period (Ref, 1986-2005). Global warming level is computed as the difference in global mean surface air temperature between EOC, MOC or BOC and either the reference period (top axes) or the pre-industrial period (P-I, 1851-1880) (lower axes). Each point corresponds . . .. . .
Warming levels of 1.5 °C and 2 °C compared to pre-industrial are shown with the vertical dashed lines. Regression lines are shown for the response at EOC, MOC, BOC or all three periods (ALL) (except for P). Mean values. . .."**

OK. This is now Figure 6. The Figure caption was changed accordingly. (p. 23)