Suggestions from Referee #1

We thank Reviewer #1 for his/her useful questions and comments on our manuscript. Please find below a detailed feedback to individual comments and questions.

Main comment:

The single main deficiency of the work is the missing validation of the SAFRAN data with respect to extreme snowfall amounts and their temporal trends. Several previous works are cited, but much more information is necessary in my opinion. Potential deficiencies of the reanalysis dataset in representating extreme snowfall woudl ultimately affect the conclusions of this work. I strongly suggest that the authors try to better motivate the use of this reanalysis dataset for their purpose.

Motivation for the SAFRAN reanalysis. The main motivation to rely on the SAFRAN reanalysis is that it provides consistent data to study the evolution of meteorological conditions and their impacts. Indeed, the direct use of observations have several drawbacks: they are intrinsically uncertain, like any observation, they often contain temporal gaps, and pose, like any point observation, representativeness challenges due to local characteristics of the location of the observations points (orography, vegetation). On the other hand, the SAFRAN reanalysis provides “augmented observations” that draw on all available observations and are supplemented with physical laws driving the numerical weather prediction model used for the reanalysis process (mass, momentum and energy conservation within the meteorological), and can reduce uncertainty by gathering strength from the combination of multiples sources of observations (temperature, precipitation). Indeed, the SAFRAN reanalysis includes precipitation in-situ data in its analysis scheme (in contrast to other popular global reanalysis products such as ERA-Interim/ERA5 reanalyses).

We further consider that we cannot “validate” SAFRAN but rather “evaluate” it, like for any geophysical product. Indeed, validating SAFRAN would imply that in situ measurements should be considered as a reference because they directly measure snowfall, and are therefore considered as the truth. However, all observations carry uncertainty and this is particularly the case for in situ snowfall measurements (see e.g. Nitu et al., 2018). Firstly, they cannot span the spatial heterogeneity, which is critical in mountainous areas (local effects due to the topography, winds, etc.). Snowfall measurements are also very uncertain for high precipitation intensities, usually associated with strong winds due to gauge undercatch issues. They are also not necessarily consistent due to the variety of measurement tools and types used in the measurement networks (automated, manual, temporal resolution).

Evaluation of the SAFRAN reanalysis. A specific evaluation of SAFRAN for extreme snowfall goes beyond the scope of this article and could easily result in a scientific publication on its own. In our study, we already mention that the “SAFRAN reanalysis has been evaluated both directly with in situ temperature and precipitation observations and in-directly with various snow depth observations”. In the revised version of the manuscript, we will add the following additional comparisons:

- *“In Vionnet et al. (2019), the SAFRAN reanalysis has been evaluated for snowfall against two numerical weather prediction (NWP) systems for the winter 2011-2012. They find that seasonal snowfall averaged over all the massifs of the French Alps reaches 546 mm in SAFRAN, 684 mm in the first NWP, and 737 mm in the second NWP. In details, they find that SAFRAN significantly differ from the two NWP systems in (i) areas of high elevation, probably due to the limited number of high-elevation stations and gauge undercatch (ii) on the windward side of the different mountain* *ranges due to the assumption of climatological homogeneity within each SAFRAN massif."*

- *"In Menegoz et al. (2020), the SAFRAN reanalysis has been compared to the regional climate model MAR which uses ERA-20C as forcing. They found that the vertical gradient of annual mean of total precipitation of SAFRAN is generally smaller than those simulated by MAR."*

Therefore, the SAFRAN reanalysis most likely underestimates high-elevation precipitation (above 2000 m), which probably leads to an underestimation of high-elevation snowfall. This deficiency does not affect the main conclusion of our article, with a majority of decreasing trend below 2000 m, and of a majority of increasing trend above 2000 m. However, this deficiency affects the value of extreme snowfall, i.e. 100-year return level (Fig. 9) and the scale of changes of 100-year return level (Fig. 8), which may be underestimated above 2000 m.

Minor comments:

Line 78-80: This section is far too short, a much better motivation for using SAFRAN instead of, for instance, station series, needs to be provided (see above). A dedicated validation exercise would help.

For an answer to this suggestion, we refer to our answer to the Main comment.

Chapter 4.1 and further: I'm not sure if the term "pointwise" should be used here. I understand the meaning and the difference to "piecewise" in statistical terms, but pointwise could be misunderstood as being based on observations taken at individual locations/points (which is not the case here)

We understand that these terms could be misunderstood. Thus, in the revised manuscript, we will add the two following sentences in In Sect. 3.4 to clarify this fact.

- *"Pointwise distribution stands for a distribution fitted on the annual maxima from a single elevation of some massif."*

- *"Piecewise models stands for models fitted on the annual maxima from a range of elevation of some massif."*

Lines 189-199: This section is basically a very brief description of one figure after the other. It should be extended and some more information on each figure and a brief description of what they show and what this means) needs to be provided.

We agree that the description for Figure 8 is rather short. In the revised manuscript, we will add the following informations for the changes of 100-year return levels:

"*In Figure 8, we display the change of 100-year return levels between 1959 and 2019 for each range of elevations. At 500 m, we observe that eight massifs have a stationary trend, and five massifs located in the center of the French Alps have a significant decreasing trend. We also note that two massifs located in the Western French Alps have an increasing significant trend, with an absolute change of 100-year return level close to +20 kg m^{-2}. At 1500 m, six massifs in the center of the French Alps have decreasing trends. At 2500 m, we observe a spatially contrasted pattern: most decreasing trends are located in the north, while* *most increasing trends are located in the south. We discuss this pattern in Sect. 5.4. At 3000 m, we observe that the six massifs with a significant increasing trend are located in the South of the French Alps.*"

Chapter 5.1: This chapter is very important as it highlights potential limitations of the work. However, the implications of these limitations for the interpretation of the results and for the conclusions remain largely unclear. Are the conclusions valid nevertheless or do they have to be questioned?

In this section 5.1 we highlight the main hypothesis of our study:

- annual maxima from a range of elevation of some massif are assumed to have the same temporal trend,
- conditional independence of the annual maxima given the vector of parameters,
- the shape parameter is considered constant.

We do not believe that these hypotheses are limitations, but rather "classical" assumptions that increase the robustness of our conclusions by avoiding, e.g., overparameterization. Firstly, we believe that the former hypothesis is reasonable. Then, the two latter hypotheses are standard in the field of extreme value analysis, and seem a good choice given the fact that we only have 61 annual maxima for each elevation. Indeed, instead of the "conditional independence" hypothesis, an explicit modeling of dependence (with max-stable processes for instance) would lead to over-parametrized models. Moreover, a time-dependent shape parameter can distort the form of the distribution w.r.t. years, which is likely to lead to overfitting when we model changes w.r.t. 61 years.

Lines 244-245: This implication does actually only hold if the past trends would continue into the future. Do you have any evidence for this?

We will replace the potentially misleading sentence "*should ensure that protective measures can cope with this increase.*" with "*should ensure that protective measures are still valid after such an increase*."

Lines 275-277: What this basically means is that mean temperature is not the only control. This could be written much clearer.

We will add an introductory sentence to this paragraph: "*Thus, this spatial repartition of changes cannot be solely explained with the spatial repartition of mean temperature*."

References

Ménégoz, M., Valla, E., Jourdain, N., Blanchet, J., Beaumet, J., Wilhelm, B., … Anquetin, S. (2020). Contrasting seasonal changes in total and intense precipitation in the European Alps from 1903 to 2010. *Hydrology and Earth System Sciences*, 1–37. Retrieved from https://doi.org/10.5194/hess-24-5355-2020

Nitu, R., Roulet, Y. A., Wolff, M., Earle, M., Reverdin, A., Smith, C., ... & Yamashita, K. (2018). *WMO Solid Precipitation Intercomparison Experiment (SPICE)*. Tech. Rep., World Meteorological Organization, 2018. a. https://library.wmo.int/doc_num.php?explnum_id=5686

Vionnet, V., Six, D., Auger, L., Dumont, M., Lafaysse, M., Quéno, L., … Vincent, C. (2019). Sub-kilometer Precipitation Datasets for Snowpack and Glacier Modeling in Alpine Terrain. *Frontiers in Earth Science*, *7*(August), 1–21. https://doi.org/10.3389/feart.2019.00182