Suggestions from Referee #3

We thank Reviewer #3 for his/her useful questions and comments on our manuscript. Please find below detailed feedback to individual comments and questions.

Major comments:

Extreme snowfall or even snowfall in general from Safran reanalysis, at least to my knowledge, has so far never been evaluated with measurements as e.g. described in Gaume et al. (2013) or Nicolet et al. (2016). I believe it's necessary to show at least some comparisons with the available time series.

We agree that this is a critical point which needs to be further discussed (see our response to the reviewer #1).

Moreover, the following points are missing, which are both important in case of comparisons with measured snowfall water equivalents: a) Is the reanalysis able to reproduce rain-on-snow-events within a 24 h time period? b) Are the reanalysis data refer to the same 24h period as the measurements?

The SAFRAN reanalysis consists of hourly rainfall and snowfall values, so that when it is used as input to a snow cover model (e.g. the Crocus snow cover model as part of the S2M reanalysis), rain-on-snow events can be specifically analyzed. However, in the case of this particular study, we only focus on snowfall (precipitation) data hence we do not specifically analyze rain-on-snow events. The SAFRAN reanalysis system primarily operates at the daily time scale, using precipitation data from 6:00 to 6:00, which are then disaggregated to hourly time resolution as a separate part of the SAFRAN system. The daily data used here in the analysis are also integrated from 6:00 to 6:00, which minimizes any influence of the time disaggregation part of SAFRAN on daily precipitation values.

The examples of extreme snowfall or roof collapses provided are used to justify the use of the 100-year level, although all these examples are not caused by one extreme snowfall alone. Most of the examples also concern elevations below 600 m, which are not analyzed in this study at all. Please elaborate.

The 100-year return period was chosen because it is the largest return period considered in the Eurocode to build structures (Cabrera et al. 2021). We also believe that the 100-year return period is the most familiar return period for non-experts as it corresponds to a centennial event. Following the suggestion of referee #2, we will add in an Appendix, the equivalent of Figure 6 and Figure 8 for the trend in 10-year return and the 50-year return level.

We agree with Reviewer #3 that roof collapses are often not caused by one extreme snowfall event alone. As mentioned in another answer to one the suggestions of Reviewer #3, we will add the sentence "*extreme snow events are often triggered by one snowfall event but depend on other factors such as accumulated snow or wind.*"

The examples have been moved in the Perspective section of the article, where we argue that specific statistical tools are needed to study extreme snowfall in low-land areas below 600 m such as compound risk analysis, or mixed discrete-continuous distributions to handle the high number of annual maxima equal to zeros, which correspond to years without any snowfall.

The linear elevation dependence considered in L100 is in contradiction to doi.org/10.1029/2009WR007916. Please elaborate.

Indeed in paragraph 23 of Blanchet et al. (2009), there is a detailed analysis about the change in the location parameter with elevation. They find that this parameter strongly increases below 1200 m, and less strongly between 1200 m and 2200 m. This agrees with our piecewise linear elevation dependence. However, they also find that above 2200 m, the location parameter is almost constant with the elevation. Indeed, this is different with the linear dependence that we find for all the massifs (including the massifs close to the Swiss border) in our preliminary analysis (Fig. 2).

We note that the word “contradiction” seems strong, because it can be seen in Figure 3 of Blanchet et al. 2009 that the location parameter still increases slightly above 2200 m. This “contradiction” might be due to various differences in the data used (length of study periods), as well as to various climatological (differences between areas) or statistical reasons.

Above all, we noticed that Blanchet et al. (2009) actually focus on the new snow depth (snow depth accumulated in 1 day, referred to as “snowfall” in their article) while we focus on the snowfall (amount of solid precipitation in 1 day). Therefore, we believe that the elevation gradient of Blanchet et al. (2009) that we found in paragraph [31] of Gaume et al. (2013) are erroneous. Indeed, the gradient of new snow depth is reported in kg m^{-2} / 100 m in Gaume et al. (2013), while it should have been reported in cm / 100 cm. Therefore, we will remove the reference to Blanchet et al. (2009).

Minor comments:

Title: I’d suggest to extend the title with «based on reanalysis» as the used snowfall data are not measured

We prefer not to change the title so that it remains a short and concise description of this study. We already indicate in the Abstract that our study is based on a reanalysis. Furthermore, the snowfall data are strongly based on measurements, since 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 fear that adding “based on reanalysis” could be misunderstood as “not representative of real-world trends”.

Table 1: Why is Faranda et al (2020) not listed?

Faranda (2020) indeed focused on extreme snowfall (and we cite this work in the Discussion section). However, Table 1 focuses on the elevation-dependency of trends in extreme snowfall. To the best of our knowledge, Faranda (2020) rather focused on the spatial distribution of trends, but did not mention/illustrate the dependence with the elevation.

S nd : Why the mean of maximum? It’s just the annual maximum snowfall in N consecutive days.

This notation was chosen based on the Table 2 of Frei et al. (2018) where S1d denotes the mean maximum of 1-day snowfall.

L61: Does 600 m elevation mean the 300 m elevation band between 450 and 750 m? Please clarify?

For the SAFRAN reanalysis “600 m elevation” means precisely that the corresponding data is taken at 600 m of elevation.

L242: I don’t find the 1.5 kg m-2/100 m in Blanchet et al. (2009)?

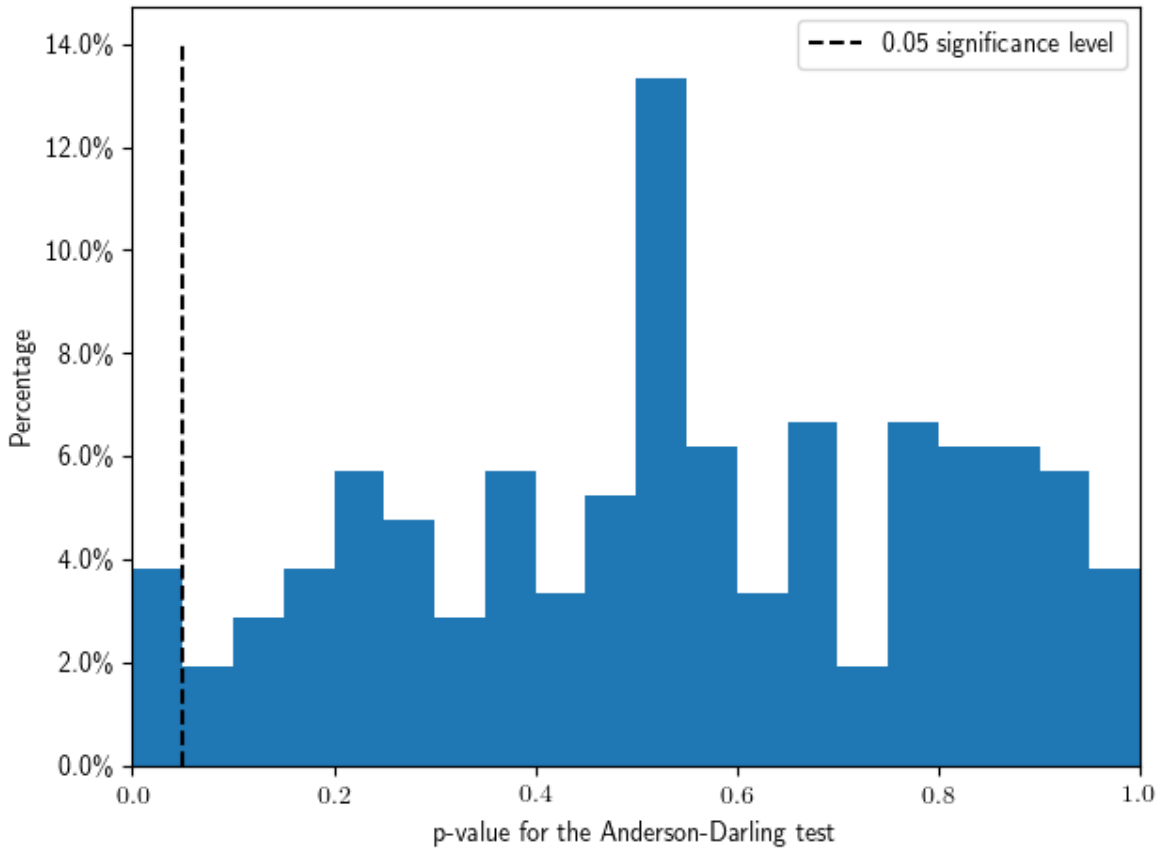
The elevation gradient for the location and scale parameters of Blanchet et al. (2009) have been found in paragraph [31] of Gaume et al. (2013). However, Gaume et al. (2013) do not mention how they obtained these values. Gaume et al. (2013) report that Blanchet et al. (2009) focus on snowfall measured in kg m^{-2} while they focus on “new snow depth” measured in cm (which is called “snowfall” in this article). Therefore, since Blanchet et al. (2009) did not study elevation gradient for snowfall, we will remove the reference to this article.

L243ff: This study is about annual maximum snowfall, which is only at low elevation a concern in regard to infrastructure. Avalanche defense structures or settlements in the mountains are endangered by annual maximum snow depth and not one snowfall event. Please rephrase!

We agree that settlements are rarely endangered by one snowfall event. Therefore, we will reformulate this section 5.3, and will add that “*extreme snow events are often triggered by one snowfall event but depend on other factors such as accumulated snow or wind*.” It is also true that individual snowfall events are particularly relevant, especially due to the fact that snow removal procedures are applied on top of buildings, so that for critical buildings there is generally no season long accumulation of snow on their roof.

L325: How is a “good” fit defined in numbers? What did you do for those cases, where no good fit could be obtained?

For a quantitative evaluation of the goodness of fit, we rely on the Anderson–Darling statistical test, which is the most powerful test for the Gumbel distribution (Abidin et al., 2012). This test assesses if the residuals follow a standard Gumbel distribution (see the Appendix A of our article for a definition of these residuals). For every selected model, the p-value of this test was computed for each elevation inside a range of elevation. On the Figure below, we observe that most tests are not rejected (with a 5% significance level) which quantitatively justify that the fits are generally “good”.



Legend: Distribution of p-values for the Anderson-Darling test. For every selected model, a p-value was computed for each elevation inside a range of elevation.

In the case where no good fit could be obtained, we would have excluded the corresponding time series. However, in practice with the Q-Q plot we did not observe any fit to exclude. This adequate goodness of fit might be due to the fact that we consider 8 simple models (Tab. 2) that can cover a wide range of behavior.

L251-252: Is there a specific reason you distinguish here below 1000 and above 3000 m, in contrast to the separation below and above 2000 m in the abstract and the conclusion?

Our analysis shows that the contrast is mainly between below and above 2000 m for the past trends. In the sentence L251-252, we speculate about projected trends. Assuming that our past trends continue into the future, and following the literature (Tab. 1), it seems reasonable to state that it is likely that extreme snowfall will, in the near future, decrease below 1000 m, and increase above 3000 m. However, for elevations between 1000 m and 3000 m, results vary in the literature (Tab. 1).

L254-262: It seems strange that you write in these lines about increase in “extreme precipitation”, “high latitude” and “moderate extreme snowfall”, which all have so far not been analyzed in this study.

We agree that this part is unnecessarily speculative. In the revised manuscript, this section will be mostly deleted. Some of these elements have been added at the end of the perspective section.

L266-274: This might be true for moderate extreme snowfall but not for the annual maximum.

We did not understand this suggestion. Between L266-274, we indeed talk about extreme snowfall but not about annual maximum. In our manuscript, extreme snowfall stands for 100-year return level, while moderate extreme snowfall stands for the mean annual maxima.

L305-307: See my comments to lines to L243ff.

We moved this sentence to paragraph 5.3 that starts with L243ff.

References

Abidin, N. Z., Adam, M. B., & Midi, H. (2012). The Goodness-of-fit Test for Gumbel Distribution: A Comparative Study. *Matematika*, *28*(1), 35–48. Retrieved from https://doi.org/10.11113/matematika.v28.n.313

Blanchet, J., Marty, C., & Lehning, M. (2009). Extreme value statistics of snowfall in the Swiss Alpine region. *Water Resources Research*, *45*(5). https://doi.org/10.1029/2009WR007916

Cabrera, A. T., Heras, M. De, Cabrera, C., & Heras, A. M. De. (2012). the Time Variable in the Calculation of Building Structures . How, 1–6. Retrieved from http://oa.upm.es/22914/1/INVE_MEM_2012_152534.pdf

Faranda, D. (2020). An attempt to explain recent trends in European snowfall extremes. *Weather Clim. Dynam.*, 445–458. Retrieved from https://doi.org/10.5194/wcd-1-445-2020

Frei, P., Kotlarski, S., Liniger, M. A., & Schar, C. (2018). Future snowfall in the Alps: projections based on the EURO-CORDEX regional climate models. *The Cryosphere*, *12*(1), 1–24. https://doi.org/10.3929/ETHZ-B-000233992

Gaume, J., Eckert, N., Chambon, G., Naaim, M., & Bel, L. (2013). Mapping extreme snowfalls in the French Alps using max-stable processes. *Water Resources Research*, *49*(2), 1079–1098. Retrieved from http://doi.wiley.com/10.1002/wrcr.20083