Reviewer’s comment on the manuscript “A non-stationary extreme value approach for climate projection ensembles: application to snow loads in the French Alps” by Erwan Le Roux et al.

General comments

In this manuscript, the authors show an application of the block maxima approach of Extreme Value Theory (EVT) to annual maxima of snow load for 23 massifs in the French Alps at 1500 m elevation, in a non-stationary setting using a multi model ensemble consisting of General Circulation Models (GCM) and Regional Climate Models (RCM). The aim is to estimate 50-years return levels for the time period 2019-2100. They use a reanalysis data set for the historical period (representing the observations), and a set of RCMs and CMIP5 GCMs. The non-stationarity of this approach relies in using piecewise linear functions linking the evolution in time of the Generalised Extreme Value (GEV) distribution parameters to a smoothed temperature anomaly w.r.t. the pre-industrial period. The estimation is done based on the information provided by the observations and all models together, thus relies on more data than what one would have by considering the models separately. The link between models and observations is based on “adjustment coefficients” which allow the connection between the block maxima statistics of the models and of the observations. The authors test different parametrisations for this adjustment: “zero adjustment”, “one for all GCM-RCM pairs”, “one for each GCM”, “one for each RCM”, “one for each GCM-RCM pair”. They choose the optimal parametrisation for each massif based on a mean logarithmic score. Similarly, a logarithmic score is used to choose the optimal number of linear pieces for each massif.

The paper raises a very important issue: in order to obtain robust estimates, especially in a non-stationary setting, we need to use a multitude of available information sources. However, it is a challenge to connect properly these different information sources and requires a lot of scientific effort. I consider this paper to be one of the initial steps in the right direction regarding the quantification of changing extreme events due to global warming. The paper has a clear structure, the language is understandable despite the many technical details and steps, the figures are illustrative. I appreciate a lot that the authors discuss critically the drawbacks of their method in Sec. 5.2, and that they provide a clear and concise overview of related studies and how this work is embedded in the current scientific literature. Another strength of this work is that a range of possible parametrisations is explored and the optimal one is chosen in an objective way, based on a mean logarithmic score. I tend to suggest the manuscript to be accepted for publication, but only after a few very important issues are clarified.

I. My main criticism is that the authors use a non-stationary GEV setting without showing that the annual maxima of snow load are properly estimated by a GEV distribution. The authors write that “…due to these theoretical justifications, the GEV distribution enables a robust estimation of return levels”. Yes, the GEV distribution enables (under certain conditions) a robust estimation, but only if

1) the chosen block size n is large enough, as the theory applies for $n \to \infty$. The block size of one year is not automatically large enough (one might need 2 years or 10 years or even longer) – it has to be shown that the annual maxima can be reliably estimated based on a GEV distribution (see convergence plots and diagnostic plots in Coles, 2001).

2) the auto-correlation is weak enough - the stronger the auto-correlation, the larger the smallest block size for which the GEV limit is valid. These convergence issues are discussed extensively in Galfi et al. (2017, Complexity).

I ask the authors to show that the convergence to the GEV distribution is good enough for the chosen block size, i.e. that the annual maxima are properly modelled by a GEV distribution. This should be done before building up the non-stationary framework. If this cannot be shown, I'm afraid that the whole experimental setup described in the paper is useless, as it is a necessary condition. In the theoretical description of the GEV distribution in Section 3.1 more emphasize is needed to underline the asymptotic nature of the theory.

III. It is a bit disappointing that even after performing the adjustments and applying the complex methodology, the results still do not follow the observations, as shown by Fig. 4 and discussed shortly in the text. One reason for this could be that, as the authors explain, the evaluation set for adjustment parametrisation contains only a few (24, 17, 10) maxima, thus the selection of the optimal parametrisation might be misleading. I believe a more thorough discussion is needed here. It is also not clear for me how to overcome the issue regarding the estimation of the shape parameter: in case of no adjustment, the model results do not follow the observations, in case it is adjusted it can lead to prediction failures. The authors do not suggest any solution for this, although it would be crucial for the applicability of the method in future studies.

IV. It is not totally clear for me why is it important to know return levels of a variable whose extremes are expected to decrease with global warming (assuming a simple direct relationship between them), thus becoming less extreme? I think the authors should put more emphasize to show the relevance of this subject.

Specific comments

L93 Are the time periods “historical” 1951-2005 and “future” 2006-2100 correct?

L98 It is not mentioned to what “Crocus” refers.

Eq.4 There should be two equations here: $\hat{\theta} = \text{argmax}(p(y|\theta))$ and $p(y|\theta) = \prod \ldots$

L177-L178 It is not totally clear what “RL50” stands for because it is written that “is computed without adjustment coefficients”, but later in Figure 4 it is used also for the case with adjustment coefficients.

L210 & caption of Fig. 4 GCM-RCM pair instead of RCM-RCM pair

L220 “adjustment” written twice

Figure 4: I think that in case of the line with the “warm” colours the legend should be different for the 4 subplots – the legend is the same for each subplot, although the text and the figure caption suggest the opposite.

Figure 5: To which year or period do these return levels refer?