This paper assesses changes in the duration of dry spells and temperatures during spells over the Pyranees for an observation period (1981-2015) and in future projections from an ensemble of regional climate models. The authors find a significant increase in temperatures during dry spells over the observation period and in projections from future climate simulations, but little change in the duration of dry spells. The paper is generally well presented but there remains many aspects that need clarification in their methods and in the presentation of their results. For instance, it is unclear in many figures what exactly has been done to produce the specified result. There is also a major issue with the use of bias correction, which is unjustified in this case, particularly with respect to bias correcting the duration of dry spells. These biases likely result from biases in the persistence of large-scale anti-cyclones which cannot be corrected simply via quantile mapping. Therefore, I unfortunately cannot recommend the paper for publication in its current state. However, I feel this research is very relevant and could be brought to publication standard. Specifically, through better discussion of the model biases, the potential sources of such biases, and the how such climate model limitations impact the confidence we can have in the future projections of such events. The reasoning behind my decision is explained in more detail below.

Comments (**P: Page, L: Line Number**):

- P3 L84: The analysis by Jacob et al. (2014) does not include any validation, only an analysis of future changes in a range of metrics in these simulations.

- Regionalisation:
    - P4 L97-105: The authors suggest that there is no variability throughout the assessed region when long dry spells occur as there is an "identical synoptic behaviour pattern throughout the region". What is the motivation is for using this regionalisation approach for the analysis of long dry spells?
    - P4 L112: what is meant by iterations in this case, and how does this ensure a robust regionalisation? Please explain more precisely.
    - P4 L113-114: What variance is it explaining? Daily temperature and precipitation?
    - P4 L113-114: How is the total explained variance calculated? What output is given from the k-means algorithm to do this? Please be more precise.

- Event definition:
    - P6 L139-140: What is meant by annually? Do you mean that you extract only one event per year?
    - P6 L140-142: Does this mean that there are multiple EM events in one dry spell? If so, do you consider all of these 'EM events' as independent events such that there would be more EM events than M or D events by definition? Please explain more precisely.

- Bias correction
    - I'm not convinced that bias correction via quantile mapping is appropriate here. It is a simple method that is used to make simple corrections to climate model output. The method only adjusts each quantile of the RCM distribution to the corresponding quantile of the observed distribution, and so it is trivial that the bias corrected distribution will be similar to the observed distribution, as is shown in Figure 6. See Maraun et al. (2017) who consider an extreme example of quantile mapping in which

the distribution of temperature from the Pacific Ocean to precipitation in central Europe. If using quantile mapping, it should be clear what is driving the bias in a given variable.

  - Quantile mapping is particularly inappropriate in the case of duration. The biases seen in duration (Figure 6) are derived from the lack of persistence of dry days in the underlying precipitation time series. This itself is driven by a lack of persistence in large-scale drivers (e.g. persistent anti-cyclones). A simple bias correction via quantile mapping of the Duration distribution cannot correct biases in the large-scale circulation and will result in fictitious events in the bias corrected distribution.
  - Furthermore, quantile mapping in this case will just hide significant and relevant uncertainties in these climate model simulations. For instance, if the models cannot represent the persistence of long dry spells, then we cannot know with any confidence how such events might change in the future. This is a reality we are faced with in the community which cannot be simply fixed via quantile mapping.
  - The uncertainty is hidden in the results obtained from the bias corrected distributions, and I would not be confident in the robustness of the results. I think it would be more informative for the authors to present the relevant biases of these models and discuss the implications of such biases for the future projections. This could help as feedback to model developers in order to improve these climate model biases.

- Figure 3: Have you taken the average of all events that exceed the local 90^{th} percentile? Is this average sensitive to the occurrence of single events? The figure seems a little noisy in places which might not be expected for metrics of such large-scale events. I'd imagine looking at the 95^{th} percentile would be more robust.

- Figure 4: The average of EM is larger than that of M by construction of the analysis, it is a trivial result. You are comparing the unconditional distribution of M with the distribution of EM which is a conditional distribution of temperature given that it exceeds the 90^{th} percentile. EM is different from M because you impose a threshold on temperature. Maybe I have missed the point but I do not see the relevance of this figure, please explain the significance of this result. Is it simply that the average temperature of dry spells with temperatures above the 90^{th} percentile are warmer than dry spells where no threshold is imposed?

- Figure 5: What were these trends calculated for? Are all events considered or just the annual maximum? If it's the former, how is the resulting slope interpreted given that there will be a different number of events each year?

- Figure 6: What are the biases calculated between? The mean of the distributions or some other metric? Please specify.

- Figure 8: What is the 7-year moving average taken of? From all events in the 7-year period? Please specify.

- Figures 10, 11, 12: This is a nice of visualising the change in the bivariate distribution. However, there are a number of aspects that need clarification:
  - Is this figure for one model only? Or do you pool the events from all models into one distribution?

- How do you compute the linear regression shown in each panel? Specifically, what values are used to construct it?
- From your definition of EM, you would obtain multiple values of EM per event. What do you plot against Duration in these figures? Is it one EM value per event? Or is each EM value considered such that the same event would be repeated multiple times in the scatter plot?

- P. 22 L337-338: It is mentioned that there is no change in duration, but the NMED and SMED regions show an increase in mean Duration for RCP8.5 in Spring and Summer (Figures 10 and 11), and it seems there are more very long duration events also from counting the number of dots in the scatter plot.