Reply to Reviewer #2 comments

This article assesses long duration dry and hot events over China. The analysis is done locally at each grid but also includes 3D analysis of the events taking their duration, temperature magnitude and contiguous area into account. The paper presents the climatology of each characteristic and assesses changes in the characteristics over the observation period. The paper is well written and the results are nicely presented. I can recommend this paper for publication after some minor revisions that I have outlined below.

Response: We thank the reviewer for the time in reviewing this manuscript. The comments are quite valuable and helpful for revising and improving our paper. We have clarified some parts of the manuscript to make it clear and concise.

The authors point out that the gridded dataset is based on a number of underlying observation stations and briefly mention at the end of the manuscript that a comparison with station data would be interesting to assess in further studies. However, I think it is important to show or discuss the spatial distribution and density of these stations. This would highlight if some regions are more observed than others and whether we should trust the changes found in regions with low observation densities. For example, in Figure 12b, is it surprising that C7 has no events for three decades (from 1990 onwards)? How trustworthy is the data in this region?

Response: Thank you for pointing this out. We agree that the spatial distribution and density of the stations may play a role in the uncertainty of gridded datasets. The gridded observation dataset we used here is based on observations from over 2,400 meteorological stations in China. Compared to the high density of stations over eastern China, the stations over western China are relatively sparse, especially over the northern part of the Tibetan Plateau and Taklimakan desert in southern Xinjiang (the location of the C7 cluster), which leads to a great uncertainty in these regions.

We have added the following text to the manuscript.

"Note that the station density in western China is lower than in eastern China, leading to a great uncertainty in this region, with the largest uncertainty over the

northern part of the Tibetan Plateau and Taklimakan desert in southern Xinjiang (Peng and Zhou, 2017; Wu and Gao, 2013)."

"... It should be noted that our results are based on gridded data, which may have uncertainties due to station density**, especially over western China with relatively few stations**."

References

Peng, D., and Zhou, T.: Why was the arid and semiarid northwest China getting wetter in the recent decades?, J. Geophys. Res.-Atmos., 122(17), 9060-9075, https://doi.org/10.1002/2016JD026424, 2017.

Wu, J., and Gao, X. J.: A gridded daily observation dataset over China region and comparison with the other datasets (in Chinese), Chinese J. Geophys., 56(4), 1102-1111, https://doi.org/10.6038/cjg20130406, 2013.

P3 L71: I suggest replacing 'meteorological drought' with a 'dry spell'. The two terms seem to be used interchangeably throughout which might become confusing for readers as 'meteorological drought' has been defined in many different ways in the literature (e.g. some define meteorological drought with SPI), while a dry spell definition is more precise and will not become confused with different definitions.

Response: Thanks for this suggestion. We concur with the reviewer and have replaced 'meteorological drought' with 'dry spell' in Line 75 to make it precise.

P3 L89: The metric 'Count' can be difficult to interpret and I don't think it's the best metric to use when assessing events defined by their duration. For example, each summer has 92 days, if you have 1 event that lasts 80 days and rain on the remaining 12 days, you have 1 event. Equally, you may also have a wet summer with one event lasting 14 days, or multiple short duration events that leads to a high count. This is an issue in interpreting changes in the metric. For instance, a reduction in the number of events could mean they are less frequent or that are more persistent. Perhaps an explanation of this would help though I see that the analysis of event counts is supplemented with changes in the number of LDDH days (Figure 3), which is helpful.

Response: We agree with the reviewer that caution must be taken when assessing events defined by their duration. Therefore, in addition to the number of such persistent extreme events, other metrics such as the number of total LDDH days and duration are also considered in our analysis.

P5 L127-128: Perhaps you could be more precise here as well as give an example of what you mean by a dominant spatial pattern. Is it finding clusters of similar events in different regions within the analysis area or the shape of the patterns (etc.)? Also, please provide the motivation behind this cluster analysis and further discussion of the relevance of the results obtained from this.

Response: Thank you for this comment. The motivation for the cluster analysis is to distinguish clusters of these events following their geographical location. These spatial clusters allow us to investigate the characteristics of the large-scale SLDDH events in different regions. Our results show that SLDDH events in China exhibit distinct characteristics in different geographical locations. For example, SLDDH events in northern China are generally more frequent and have a greater spatial extent than those in the south. The mechanisms/processes for their occurrence may be regionally dependent. It is of great interest to extend the current analysis to examine the physical mechanisms associated with the different clusters, thus supporting the adaptive management of such extreme events.

We have clarified the clustering approach in more detail in section 2.3 as follows.

"**Here we apply a hierarchical clustering algorithm (Rokach and Maimon, 2005) to identify clusters of SLDDH events over China following their geographical location and analyze their associated characteristics.**"

Section 3.2.1: This section could be improved, particularly the description of the metrics they assess and the justification for the metrics assessed in Figure 8. It's not instantly clear what exactly the authors are assessing here. My understanding is that they assess trends in the mean annual characteristics of events. This is not so informative without an assessment of the variability within seasons (e.g. larger/longer impactful events may be averaged out by smaller/shorter, and less impactful, events). Perhaps an assessment of the trend of the maximum event each year would be more informative. For instance, one could estimate a conditional trend for the area associated with the annual maximum duration event, or vice versa.

Response: Thank you for pointing this out. In previous submission, when assessing trends in different characteristics (i.e., mean duration, mean area, and magnitude) of SLDDH events, the annual maximum is considered for mean duration, whereas the annual mean values of all events is used for spatial extent and magnitude.

We agree with the reviewer that trends calculated by considering all events may be influenced by the variability of events in each year. As suggested, we have updated Fig. 8 by considering trend of the annual maximum, the text and figure caption have been rephrased accordingly.

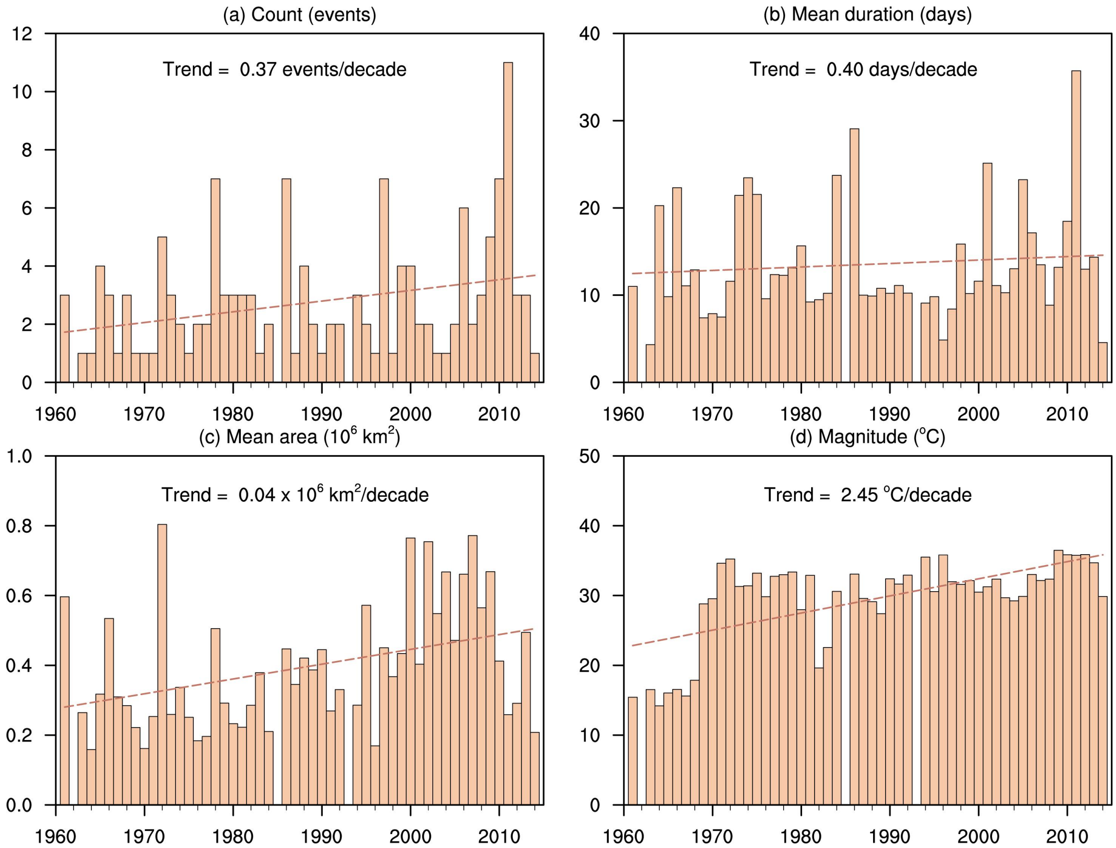

Figure 8: Time series of the SLDDH (a) count, annual maximum value of (b) mean duration, (c) mean area, and (d) magnitude over 1961-2014. The dashed line shows the linear trend.

Figure 7: It would be interesting to included magnitude in this figure also. Do larger area events have higher temperatures?

Response: Thank you for this comment. As shown in Fig. A1, the magnitude and spatial extent are not strongly correlated with each other. The events with a large areal extent can have moderate temperatures.

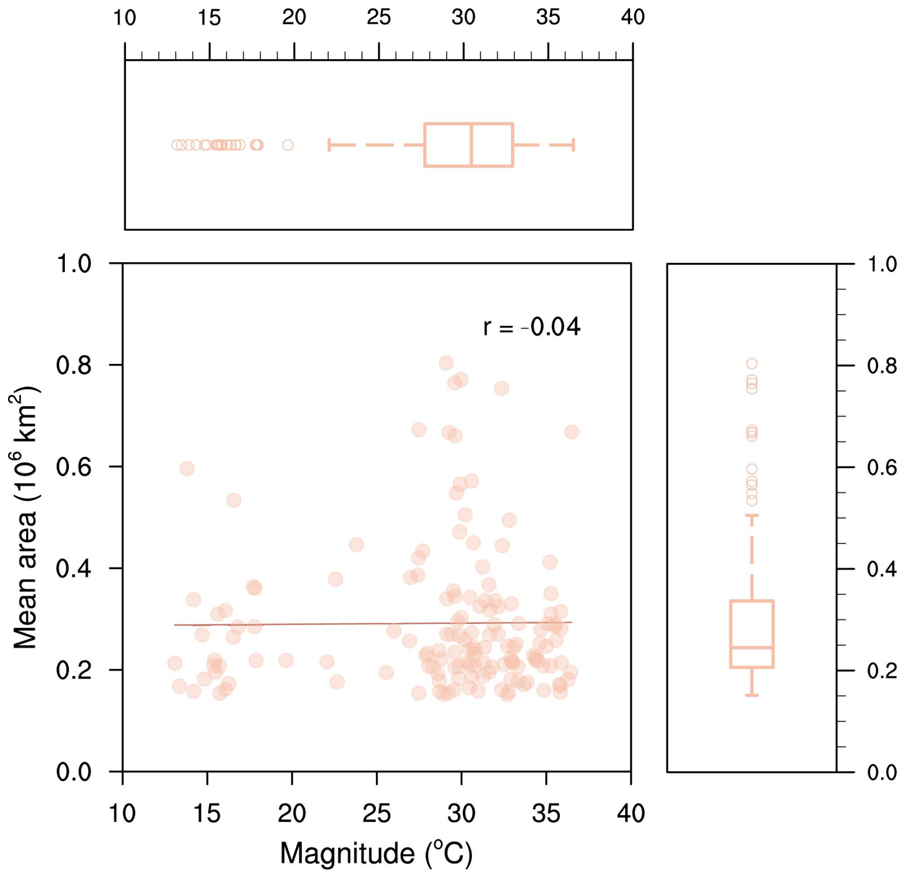

Figure A1: Scatter plot of magnitude and mean area for each regional compound long-duration dry and hot event during 1961-2014. The red line is a least squares line of best fit. Spearman correlation coefficient ($r = -0.04$, $p > 0.05$) is shown in the figure. The upper and right panels show the boxplot of magnitude and mean area, respectively.

P14 L294-5: 'The annual maximum mean duration' is confusing. I'm not sure what this means exactly, it should be clarified.

Response: The SLDDH events are characterized by their mean impacted area, mean duration and temperature magnitude. For each SLDDH event, mean duration is defined as the average duration of all cells contributing to the 3D event. In Figure 8, the annual maximum of this metric is considered. This sentence has been rephrased as "The annual maximum value of mean duration" in the revised manuscript.

Figure 12a: For events starting in August, do they automatically end at the end of August or are days in September counted also if the event persists past the end of August? If not, this might contribute to the low number of events in August.

Response: Thank you for pointing this out. Here we only consider the days during summer from June to August (92 days in each summer). We agree with the reviewer that the cutoff at the 31st of August might have an impact on event detection, contributing to the low number of events in August. To test the sensitivity of cutoff on seasonality, two cases are considered. Case 1 only considers the days during summer of JJA, while Case 2 also includes the events overlapping these months that begin before or end after this period. Thus, the events that start in August and extend beyond this month are also counted in Case 2.

Generally, the number of events begin in August is low in both cases (Table A1), although 20% of events begin in May with a mean beginning date of May 24th (all of them ends in June) in Case 2. When considering the seasonality for different clusters, the results of the two cases are similar, so this choice does not affect the overall results of the study (Figure A1). An exception is found for C5 over Northern Xinjiang, where most events begin in August (July) in Case 2 (1), which corresponding to a total of 6 events in both cases.

Table A1. The percentage of events starting in each month (%).

Month	**4**	**5**	**6**	**7**	**8**
Case 1	0	0	45.9	36.3	17.8
Case 2	1.3	20	27.3	30.7	20.7

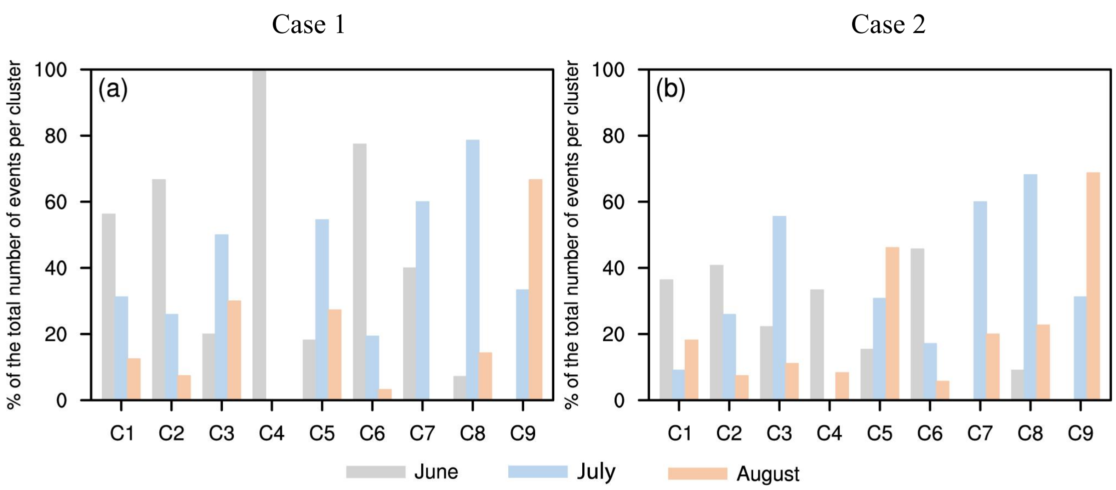

Figure A2: The percentage of events starting during June, July, and August for each cluster in (a) Case 1 and (b) Case 2. Case 1 only considers the days during summer of JJA, while Case 2 also includes the events overlapping these months that begin before or end after this period.

This has been discussed in the manuscript as follows.

"... There is a clear seasonality of the SLDDH events, with most events (45.9%) begining in June (Fig. 12a). **Note that the events are identified during summer from June to August. This might contribute to the low number of events in August. We test the sensitivity of our results to the cutoff by comparison with the results considering the events overlapping these months that end after this period. We find that the seasonality of SLDDH events are robust to the cutoff days of the analysis period (not shown).** "

Finally, we appreciate all of your insightful comments. Thank you for taking the time to help us improve the paper. We are hopeful that our revised version rises to your expectations.