We thank referee #2 for the additional suggestions to improve the interpretability of the manuscript. Referee comments are written in italics, while our responses are in bold. Lines in brackets refer to latest tracked-changes version.

---

1. I appreciate the authors' clarifications in the introduction. I would recommend a further edit on l. 51 when the authors refer to "exceptionally warm temperatures". As the authors specify in their reply, events of a similar intensity to those that they detect have previously been referred to as "moderate hot spells". They later refer in their replies to "sufficiently moderate hot days". "Exceptional" and "moderate" have very different connotations, and in the case of many of the events that the authors select, it would be more appropriate to refer to them as moderate rather than as exceptional.

(l. 44) Removed "exceptional" - has been changed to "sustained warm temperatures".

---

2. ll. 409-411 I appreciate the addition here, but this means that the sentence on ll. 410-411 now refers to the new statement "A driver need not necessarily be present …", while I assume that the authors would want it to refer to the sentence on ll. 408-409. It would be appropriate to also specify here that, given this variability in drivers and in their links to surface hot events, the persistence of a driver does not a priori guarantee the persistence of the associated surface hot event.

(ll. 398-9) We thank the reviewer for pointing out that inconsistency. The sentence at ll. 410-411 was swapped up a place, now appearing before the new statement "A driver need not necessarily be present …". Regarding the reviewer's latter comment, we appreciate the suggestion, but we are hesitant to make this claim explicitly, as our analysis does not provide direct evidence of cases where a driver persists while the associated hot spell does not. Nonetheless, we agree that this is a valuable consideration and warrants further investigation in future work.

---

3. Analysis time period: I understand the issue faced by the authors. In this case I would recommend being open about the reason for the choice of time period in the text.

(ll. 105-106) We have inserted a brief sentence in this regard.

---

4. Comment regarding non-overlapping periods. As for the previous comment, in the interest of interpretability of the methodological approach, it would be helpful to add a sentence to the paper explaining this.

(ll. 130-1) We have inserted a sentence explaining the choice of using non-overlapping periods.

---

5. I do not follow the explanation regarding why the authors select a threshold in units of standard deviations for the hot spells, but used a percentile threshold for the regionalisation. A standard deviation threshold implicitly adapts to different distributional forms and will correspond to different percentiles for different distributions. Why is this a problem for the regionalisation but not for the selection of hot spells?

We thank the reviewer for raising this apparent inconsistency. We would like to amend our original reply on this matter by stating that the regionalisation was performed in a first step, for which we used a threshold in percentiles. Subsequently, when it came to detect events, after trying multiple thresholds of both kinds (percentiles and standard deviation), we settled on one that would somewhat increase the number of cases within the reanalysis period that was acquired. This is the main reason, which was not motivated by other considerations/assumptions regarding distributions.
As the reviewer correctly points out, a threshold in standard deviation results in different percentiles for different distributions - this might ensure comparability across regions in terms of departure from local variability. In contrast, a threshold in percentiles would ensure comparability in terms of event rarity, but not necessarily in terms of how anomalous the values are relative to local variability - i.e. the 85th percentile would lead to detecting 15% of observed values at every location. That said, given that the +1σ threshold for the different regions happens to correspond approximately to the 85th percentile, the practical difference between the two approaches is minimal. Such considerations, however, did not guide the chosen methodology.

---

6. Related to the above: I again emphasise the importance of the methods not only being reproducible but also interpretable, and I would recommend that the authors add a sentence to explain their choice in the main text.

(ll. 161-2) The sentence now explicitly mentions the main reason for opting for +1σ as a threshold. As such, we have transparently explained the method and choices for reproducibility. Other researchers can recreate our results or choose different thresholds if they wish to.

---

7. Original comment on l. 158. The authors are again citing papers that do not analyse impacts, but focus exclusively on hazards. The sentence that they are supporting with these citations starts with: "From an impacts perspective", and touches on both weather persistence and impacts. One or two studies looking explicitly at impacts are required to support this – in addition to the already cited studies on persistent events.

(l. 170) We now also cite the following papers: Quandt et al. (2017), using heat indices, show that the persistent weather conditions during the 2010 Russian heatwave were indeed impactful. Though not with regard to heatwaves, Barton et al. (2016) show that the serial clustering (flow recurrence) of extratropical cyclones producing extreme precipitation have large impacts on flooding.

---

8. L. 210 I would state explicitly that the stippling should not be interpreted statistically as corresponding to a specified significance level. The current formulation referring to "the same way" does not explain the core point here.

(ll. 196-202) The sentence has been amended for improved clarification.

---

9. Choice of regions: the authors mention "a few additional long spells" in their replies yet write "average nearly twice as many long spells" in the paper. I would not equate "nearly twice as many" to as "a few", and would ask the authors to reconsider the reply they have provided to my earlier comment.

We appreciate the reviewer's interest in the results for other clusters and agree that interesting insights can be gained by including other areas. However, including results for more regions would make this already very long paper even longer. We therefore needed to choose a selection of results to present in the paper. We chose two cluster areas that show interesting differences based on the different latitudinal positions of the clusters and the different soil moisture regimes. Part of our choice is also governed by the fact that we live in some of these areas.

---

10. ll. 353 and following. I appreciate this addition, but am unclear as to why the size of the blocking system should reflect its quasi-stationarity (ll. 358). I would recommend a clarification in this respect.

We thank the reviewer for pointing this out. We have clarified the physical link between the size of the blocking system and its quasi-stationarity by elaborating on the dispersion relation for barotropic Rossby plane-wave perturbations at the beginning of said paragraph (ll. 338-340). Larger-scale systems correspond to lower wavenumbers, which are associated with slower intrinsic phase speeds. In regions of weak mean flow, such systems can become quasi-stationary (Hoskins and Ambrizzi 1993).