

We thank referee #1 for their review and various suggestions to improve the manuscript. In the following we will respond to the comments listed by the referee, which are written in italics. Our responses are written in bold.

Minor comments:

1. *L121: is there a reason for a 3-week, non-overlapping period? How does this influence your results, in particular the short heat waves?*

The choice of 3-week, non-overlapping periods has two motivations:
First, we choose 3-week periods to focus on the subseasonal timescale in which we require the co-occurrence of temperature extremes to be reflected in the resulting clusters. For the same level of event synchronicity (12.5%), clusters resulting from 3-week periods are larger and fewer than with 2- or 1-week periods. This choice is made somewhat redundant by the fact that region size and number can additionally be subjectively changed by altering other parameters such as the threshold for the dendrogram truncation or even for the binarisation before the clustering (see Section 1 of the Supplementary Materials). In the end, the choice for the parameters was tuned to end up with regions of size and location that best matched European heatwave research, whilst still being objective in its grouping of grid cells and demarcation of clusters based on observed extreme temperature variability. Long-lasting or short events can be defined for regions of most sizes. That said, one could speculate that region size would affect the distribution of event durations. In particular, very short heatwaves (1-2 days) probably affect an even smaller fraction of the regions.
Second, we choose non-overlapping periods to minimise the statistical dependence between different periods, which would affect the statistical significance testing we perform in our analyses.

2. *L146: how do you define a heat wave day? Is there a minimum spatial threshold or is one grid point sufficient?*

We define a hot day as a day exceeding the +1 standard deviation threshold of the region-averaged standardised anomalies, which were previously deseasonalised and detrended. We mention this threshold at L150. As the other reviewer pointed out, this is an important sentence that seems rather casually mentioned, so we will give it more prominence in the paragraph by mentioning it at the beginning.

The threshold is based on the standard deviation of the time series resulting from the spatial average of the standardised anomalies in the region. The exceedance was not defined based on a minimal spatial requirement. However, we did explore other methods to define a hot day and how these affect the distribution of hot spell durations, which are briefly discussed in Section 2 of the Supplementary Materials.

3. L161: *for long-lasting heat spells: do you use the same 4-5 day period for all meteorological fields?*

We use a different random seed for each meteorological field, meaning that we subsample different subperiods each time. Even with this approach, we observe a consistent correspondence between the composites across variables (i.e. in almost all cases. During the analysis, we also subsampled different periods for the same meteorological field, and this did not alter the qualitative patterns or the statistically significant areas.

4. *Figure 1: there appears to be a typo in the box next to I. (I_4^* and I_3^* must be switched)*

We understand how this may look like an oversight. In the logic of the schematic, the subspell period I_3 was selected prior to I_4 . In other words, the sequence should not matter for the labelling as the subperiods are sampled randomly across all long spells. However, to avoid confusion, we will correct this, especially since the rest of the subspells appear to be ordered.

5. *L210: the reader may be interested in the other regions too (maybe add some results to the supplementary material if it adds an added value to the study)*

We have already included the same figures from Sections 3.2-3 of the main manuscript for the other regions in Section 4 of the Supplementary Materials, for those interested in reviewing them.

6. *L259: parenthesis missing in the quotation*

Added.

7. *L286: remove point before Rossby wave*

Removed.

8. *L331: cutoffs > cutoff*

Corrected.

9. L361: *insert blank space before Wehrli et al.*

Added.

10. Figure 9c/10c: ≥ 2 instead of $2 \geq$

Corrected.

11. Figure 9/10: *the size of the squares ... —> does it mean that e.g. 66% of the heat wave it is dominated by a feature (e.g. a block)*

Yes, that is the correct way to interpret the square size.

12. L393: *remove typo before Rossby waves (WN)*

‘WN’ is meant to come before ‘Rossby wave’, as it refers to the latter’s wavenumber. This part of the sentence is meant to be read as: “Lower WN Rossby waves are more stationary than higher WN waves and can thereby enhance the persistence of surface weather systems.” The sentence will be rewritten in this way for clarity. Also, we will make sure to be more consistent with the use of acronyms.