Review of "How atmospheric and surface conditions shape recurrent heatwaves in Europe: From local to continental scales"

The work of Dillerup et al. investigates the recurrence of heatwaves in Europe using the ERA5 reanalysis during the period 1950-2024. The authors look at the recurrence at local, object and continental scales and investigate the difference between recurrent and non recurrent heatwaves, with regards to their length, intensity, spatial extent, and links with soil moisture and large scale circulation.

While the paper has a clearly defined research question that is of interest for the scientific community working on heat extremes, I found the paper confusing in how statistical and technical choices are made and the results presented, to the point that I have doubts in the robustness of what is inferred by the authors. Moreover, the authors do not really investigate or test any physical hypothesis for explaining why and how the recurrence occurs, at any spatial scale, which makes this work only descriptive. This comes as a surprise given that there is an emerging literature on the recurrence of weather systems and the feedbacks with land surfaces explaining the dynamics of persistent and/or recurrent heat events, that the authors seem to ignore. As a consequence, at the moment I am not convinced about the strength and the validity of the results presented here and I cannot recommend it for publication. Please find below my major and minor comments.

Major comments:

1. I must say that I had some difficulties understanding the different "recurrence" between the definitions at local, object and continental scales, that do not always seem to be mutually coherent to me. I wonder whether using the same word for things that measure completely different mechanisms is a good choice in terms of clarity and understanding. More precisely:
    a. Local recurrence: this definition is the easiest to follow and makes sense to me and is the closest to the current literature, although I was not sure whether when one heatwave is followed by another one the entire period was flagged as being recurrent or only the second heatwave.
    b. Object recurrence: I had a really hard time understanding the definition of recurrence here. The upper and lower quartiles of percentage of grid points in a local recurrence seem to be intrinsic properties of each heatwave object through its lifetime. But then how can heatwave objects be selected based on this upper quartile to be strongly recurrent heatwave objects ? (L174-175)
    c. Continental recurrence: this definition at the continental scale is slightly strange because it can flag as "strongly recurrent" summers with very different dynamics, for example with a recurrence of weather systems occurring in western vs eastern Europe. What is the need to aggregate a measure of summer recurrence at the continental scale ? Why can't it be done at the grid point level ?
2. Sensitivity of the separation between single and recurrent heatwaves: the authors are aware that their separation between single and recurrent heatwaves is sensitive to the arbitrary thresholds proposed, and provide some sensitivity tests. Nevertheless, I found those tests quite minimal: it does not require a lot of variability to misclassify a single heatwave to a recurrent heatwave or vice versa. Shouldn't there be a more

stringent criteria to separate the two ? Like for example at least 3-4 days below the quantile ? How does the choice of the 90th quantile affect your conclusions ?

3. Lack of a null hypothesis for the recurrence: where I struggle to make sense of the results is that I do not find any null hypothesis for how single and recurrent heatwaves should behave and thus what is surprising/unexpected in the results presented. If you consider a red noise process over the summer, and apply your selection criteria, you will find both single and recurrent heatwaves: would their statistics be different from the one you find ? In other words, can the differences observed be attributed to statistical noise ?
4. Strong spatial differences in Europe: a lot of the results presented by the authors are heatwave objects and/or metrics aggregated at the European level (e.g. Figure 3, 5, 6 and 7) without taking into account that between south Spain and Latvia the dynamics of heatwaves is pretty different. Just as an example, there are substantial differences in the standard deviation of the daily maximum temperatures between months and regions: a +1K anomaly does not always mean the same thing. Moreover, because of spatial disparities between the events identified (e.g. Figure 4), this introduces a spatial bias when averaging, implicitly favoring certain regions over others, reinforcing this effect. As a consequence, I do not know how to interpret the results shown in most figures.
5. I think the authors may find it interesting to compare their results to several works that have investigated the dynamics of long heatwaves, hot summers and/or persistent heat extremes in general, especially with regards to the question of recurrence (of heatwaves or other weather systems). Among others:

Ragone, F., Wouters, J., & Bouchet, F. (2018). Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, *115*(1), 24-29.

Ragone, F., & Bouchet, F. (2021). Rare event algorithm study of extreme warm summers and heatwaves over Europe. *Geophysical Research Letters*, *48*(12), e2020GL091197.

Röthlisberger, M., & Martius, O. (2019). Quantifying the local effect of Northern Hemisphere atmospheric blocks on the persistence of summer hot and dry spells. *Geophysical Research Letters*, *46*(16), 10101-10111.

Röthlisberger, M., Frossard, L., Bosart, L. F., Keyser, D., & Martius, O. (2019). Recurrent synoptic-scale Rossby wave patterns and their effect on the persistence of cold and hot spells. *Journal of Climate*, *32*(11), 3207-3226.

Röthlisberger, M., Sprenger, M., Flaounas, E., Beyerle, U., & Wernli, H. (2020). The substructure of extremely hot summers in the Northern Hemisphere. *Weather and Climate Dynamics*, *1*(1), 45-62.

Ali, S. M., Martius, O., & Röthlisberger, M. (2021). Recurrent Rossby wave packets modulate the persistence of dry and wet spells across the globe. *Geophysical Research Letters*, *48*(5), e2020GL091452.

Ali, S. M., Röthlisberger, M., Parker, T., Kornhuber, K., & Martius, O. (2022). Recurrent Rossby waves and south-eastern Australian heatwaves. *Weather and climate dynamics*, *3*(4), 1139-1156.

Noyelle, R., Caubel, A., Meurdesoif, Y., Faranda, D., & Yiou, P. (2025). Evolution of the dynamics of centennial hot summers in Western Europe with climate change. *Geophysical Research Letters*, *52*(14), e2025GL115552.

Noyelle, R., Caubel, A., Meurdesoif, Y., Yiou, P., & Faranda, D. (2025). Statistical and Dynamical Aspects of Extremely Hot Summers in Western Europe Sampled with a Rare Event Algorithm. *Journal of Climate*, *38*(18), 4763-4787.

Pappert, D., Tuel, A., Coumou, D., Vrac, M., & Martius, O. (2025). Long vs. short: understanding the dynamics of persistent summer hot spells in Europe. *Weather and Climate Dynamics*, *6*(3), 769-788.

Pappert, D., & Martius, O. (2026). A tale of two hot spells: dominant circulation patterns favouring persistent high temperatures in western Europe. *Environmental Research Letters*, *21*(13), 134003.

Pfleiderer, P., Noyelle, R., & Sippel, S. (2026). Heat extremes within hot summer seasons intensify more than in average summers under global warming. *Earth's Future*, *14*(8), e2026EF008241.

Minor comments:

1. L1: “Their recurrence, defined as sequences of heatwaves separated by short breaks, is projected to increase disproportionately under climate change, yet remains poorly understood”, same L42 -> my understanding of the references cited in the introduction to justify this claim (Baldwin et al. 2019 and Zhang et al. 2025) is that the two studies do not detrend the data they use to compute this metric (contrary to what the authors do), which mechanically increases the number of sequences of heatwaves. Baldwin et al. (2019) in particular write the following: “given there are more hot days, those hot days occur closer together in time and are more likely to compound.”
2. L30: “frequently associated with persistent anticyclonic circulation patterns” -> this is quite an understatement: in Europe anticyclonic conditions are a necessary condition for surface heat extremes.
3. Could the authors comment on the quality of ERA5 for soil moisture and daily maximum temperature ?
4. L107: “we first detrend daily maximum temperatures in each grid cell using a second-order polynomial" -> I am not sure this is the correct way to compute anomalies to correct for the non-stationarity of the time series. Regressing on a measure of the global mean surface temperature or regional mean surface temperature would likely be better. As you use daily maximum temperatures, have you evaluated whether the aerosols impact in the 1950-1980 period is small enough that there is not an important change in the variance of this variable ? This would bias your detection of heatwave objects.

5. It took me some time to understand over which months the heatwaves are computed, probably worth mentioning it more clearly in the method section. Moreover, is it relevant to group together months from May to September ? In May soils are usually much moister than in July and one could wonder whether one can expect a similar dynamics of heatwaves at these two moments of the year.
6. L108: “90th percentile of the detrended daily maximum anomalies for each grid cell using a 31-day running window over the 1991–2020 reference period” -> why only 1991-2020 ? Why not the entire period of study ?
7. Figure 3 and L198: “Subsequent recurrent heatwaves consistently show similar or lower anomalies than their preceding heatwaves.” -> I must say that I have a really hard time seeing any significant (in the sense of important) differences in the statistics presented in Figure 3, apart maybe for soil moisture. First, the authors say that they tested the statistical significance using a Mann-Whitney test: this test measures whether two distributions are different, whereas I think here what the authors want to measure is more whether the average or the median of the distributions are different (based on their comments in the text). Second, the differences that I read on the graph of the median of the boxplots are minuscule compared to the variability between events (visually at least a factor 10): these differences may well be statistically significant because you have a very large sample, it does not mean that they are physically significant in the sense of showing a major difference between the different events.
8. Table 1: Those results are interesting but difficult to compare to each other, wouldn’t it be possible to present them in a visually more intuitive way ?
9. L243: “Strongly recurrent heatwave objects have longer mean durations (27.91 days) than weakly recurrent ones (8.58 days), indicating that longer-lasting objects are more likely to shift spatially and subsequently return to some of the previously affected regions” -> I have the impression that this result is a tautology: don’t you define strongly recurrent heatwave objects precisely to measure that ? How is this surprising ?
10. Figure 4: I cannot make sense of the values that I read on those plots: how can there be only 20+10 heatwave objects for strongly and weakly recurrent heatwaves over the May-September 1950-2024 period for, for example, grid points in central Europe ? Does the criteria to select strongly and weakly recurrent heatwave objects filter that strongly the events ?
11. Figure 6: the differences in the attribution of weather regimes seem really small to me, probably because you mix events coming from very different places as discussed above. Is there a way to visualize the typical variability of the attribution frequency and to assess whether the differences are significant or can arise from pure variability ?
12. L363: “continental heatwave recurrence represents a distinct feature of summer rather than merely reflecting particularly hot seasons” -> I think it is statistically not possible to make this distinction, whatever the results of the authors. This comes from the fact that they study a total of 75 continental summers, from which only something like 7-8 can be considered extreme (using the same percentile as the authors use to define heatwaves). From this very low number of extreme summers, and because of the spatial variability outlined above, there is simply too much statistical noise to make the distinction claimed by the authors.