Review egusphere-2026-4179

The manuscript presents a comprehensive analysis of the June 2026 heatwave in Germany. Overall, I find this a solid study. The manuscript addresses several aspects of this anomalous heat event, combining its spatial and temporal characterization, long-term context, and circulation-based analysis. Although the study is primarily diagnostic, its originality lies in the exceptional nature of the event itself, and its characterization provides useful information for improving our understanding of such extreme events and, as the authors mention, potentially contributing to their anticipation and preparedness. However, several methodological aspects require clarification before acceptance.

General comments

1.
I am not fully clear on how the specific days associated with the June 2026 heatwave were defined. The manuscript identifies a five-day period (25–29 June), and some of the event characterization focuses on the three most intense days. However, in the introduction, this heatwave is introduced as with the WMO definition or DWD (which, please note, is not referred to in the references). Different institutions may use different definitions of heatwaves. It would be useful to clarify how these specific dates were selected and how they relate to the heatwave definition applied in the study. Are these dates identified similarly with these definitions?
Related to this point, the Methods state that the calendar-day P90 threshold is computed “at each grid point (or for each area-averaged regional series).” Are heatwaves independently identified at each grid point and separately from the regional-mean series? If so, these approaches could lead to different characteristics, right? Ideally, heatwave occurrence would first be identified locally at individual grid points and subsequently spatially aggregated, but other approaches are also possible. Please clarify exactly how heatwave events were defined for the station, gridded data and regional Germany federal states.

2.
It is not always clear which observational dataset is used for each part of the analysis. My understanding is that the long-term station data are mainly used in Section 3.3, whereas most other analyses rely on the 1-km HYRAS-DE gridded product. I suggest explicitly stating in the Methods which dataset is used for each analysis and why.
This is particularly relevant for record-breaking temperatures, since extreme values can be sensitive to spatial interpolation and area averaging. Have the authors compared the HYRAS-DE values with the corresponding station observations during the 2026 event? Such a comparison, even if presented briefly or in the Supplement, would help demonstrate that the gridded dataset adequately captures the magnitude of the observed extremes.
Relatedly, it may be worth explaining why station observations are not used more extensively in the event-level analysis, given that they provide the direct measurements underlying the record-breaking temperatures.

3.

I understand the practical reason for restricting the analysis to May–June. However, some additional climatological context would be useful.
How common are heatwaves in May–June in Germany compared with July–August? Is June already a climatologically important heatwave season, or is the timing of this event itself unusual? The manuscript mentions that early-season heat can have particularly important impacts, but a brief quantitative or literature context regarding the climatological frequency and intensity of May–June heatwaves would strengthen this argument. It would be interesting if the authors make clear how representative this restricted early-summer period is relative to the full summer season.

4.
I suggest providing a concise table summarizing the heatwave metrics used in the manuscript, including their definitions and units. This would make the methodology considerably easier to follow.
In particular, I would appreciate further justification for the definition of cumulative intensity. Since heatwaves themselves are defined relative to the calendar-day P90 threshold, I wonder why cumulative intensity is calculated relative to the climatological mean rather than relative to the same P90 threshold.

5.
In Figures 10 and 12, the dynamic component is represented using the mean T850 anomaly of the selected analogues. It would be very informative to also show, or at least quantify, the distribution of T850 anomalies among the individual analogues. The mean alone may conceal substantial variability among analogue events. Showing, for example, boxplots, violin plots, or the percentile range of the analogue temperatures would allow the reader to assess where the 2026 event lies within the historical analogue distribution. This would be particularly useful when comparing the early and late epochs.
Such an analysis could provide a more direct answer to the question of whether the 2026 event lies completely outside the range of temperatures historically associated with comparable circulation patterns, rather than only showing its difference from their mean.

6.
It is not entirely clear how the epoch-based composites in Figures 12–13 are constructed. Section 2.4 states that, for each event day, the 25 closest Z500 analogues are selected and that the analogue selection is subsequently restricted to the early (1970–2000) and late (2001–2025) periods. However, the Results refer to substantially different numbers of "qualifying analogues" in the two epochs (e.g., n = 38 and n = 87), while Figure 13 appears to show yet another pair of sample sizes. Please clarify.

There also appears to be an inconsistency in Figure 13 between the sample sizes reported in the figure panels and those given in the caption/text. Please check and correct these values.

Minor comments

L 128–130. The expression *"percentile-based algorithm equivalent to the widely used framework..."* reads somewhat awkwardly. I suggest rewriting it.

L124 I would suggest using *“early-summer period”* rather than *“short summer season May–June”*.

Figure 1. It would be useful to include the locations of the 18 long-term stations listed in Table 2. A small elevation/topography inset or background could also facilitate the interpretation of the station distribution, particularly for the mountain stations discussed later in the manuscript.

Figure 4. Please clarify exactly how the P90 shown in this figure is obtained. Is it calculated from the regional-average series? This is not clear from the caption.

Figure 12 / L351–355. The manuscript states that *“the late-epoch analogue composite is systematically warmer than the early-epoch composite on all event days.”* This is not immediately evident from Figure 12; for several days, the early-epoch values appear comparable to or even slightly higher than the late-epoch values. Please verify this statement and either provide the corresponding numerical values or revise the interpretation accordingly.

Figure 13. The figure indicates that both epochs are associated with a broadly similar blocking configuration. However, the Z500 anomaly patterns are not strictly identical and show some differences in position and amplitude. Therefore, the statement that the two circulation patterns are *“nearly indistinguishable”* may be somewhat stronger than what is directly supported by the figure.