Response to Reviewer 2

We thank the reviewer for the positive assessment and the detailed and constructive comments, which will help us to improve our manuscript. Reviewer comments are shown in italics; our responses follow in plain text.

This manuscript provides a comprehensive statistical and synoptic analysis of the June 2026 heatwave. Daily maximum temperature during May and June of 2026 is compared to the historical record. Additionally, the overall trend of heatwaves across Germany is assessed. Lastly, the flow-analog approach is used to disentangle the dynamic and thermodynamic contribution of the June 2026 heat extreme. To this end the article employs observational records, the HYRAS_DE product and the ERA5 reanalysis data set.
The work provides a timely analysis of a meteorological extreme that is unprecedented for several common heatwave metrics. Analyzing this event can help the wider scientific community to better understand the dynamic and thermodynamic effects of climate change. However, I have significant reservations regarding the choice of the May-June reference period, the chapter 3.4, the quality of the flow-analogues, and the phrasing of certain conclusions. Nonetheless, I believe that the results of this study can relevant and valuable, if the following points are addressed:

General comment 1: *The choice of May-June as reference period seems arbitrary to me. Specifically, because the event peak occurred just at the end of the reference period. The Figure 5 b)-c) indicate that two separate heat wave events have occurred in succession (at least regionally) and that shifting the start of the reference period would significantly alter the extremeness of the event. Nonetheless, I do not think that completely rewriting the study into an event-based analysis is necessary. However, a more in-depth discussion of the rational when choosing the reference period is needed. I recognize that the conclusions are tailored to the May-June reference period. Nevertheless, a sensitivity analysis with varying start dates of the reference period can put the extremeness of the heatwave into context. This should be done, at least for the Germany-wide metrics, such as the standard exceedance and the rankings of Fig. 6.*

> **Response:** The aim of the study is to characterize early-onset heat, i.e., heat occurring before the climatological peak of summer. In the revised version of the manuscript, we will motivate this choice explicitly and quantitatively: between 1951 and 2025, the annual maximum of the Germany-average TX occurred in July or August in 58 of 75 years and in June in only 13 years, and only 36 of the 255 days with Germany-average TX above 30 °C occurred in June (217 in July–August). Early-summer heatwaves have nevertheless become more frequent (years with at least one May–June heatwave: 18% in 1951–1990, 34% in 1991–2025). As suggested, we will also perform a sensitivity analysis for the Germany-wide metrics, with varying the start date (1 May, 15 May, 1 June, 15 June) and extending the window to 15 and 31 July.

General comment 2: On a related note, it should be made abundantly clear, when the May-June reference period is used and when the analysis uses data from the entire year. As I understand, in chapter 3.2 – 3.3 the May-June reference period is used and in chapter 3.4 the entire annual data is used. However, I am unsure whether this is even correct.

> **Response:** All heatwave statistics, ranks, standardized exceedances and trends in Sects. 3.2–3.4, including the trend maps in Fig. 8, refer to May–June; the calendar-day thresholds are computed from the full-year record (±15-day window), so that the thresholds at the season edges are unbiased. The circulation analyses (Sects. 3.5–3.6) use daily ERA5 fields around the event. This will be clearly stated in the revised version of the manuscript, and all metrics will be listed in a new Table.

General comment 3: *The first part of chapter 3.4 is lacking context, on how it fits into the wider field of research. Additionally, the spatial average trend values are missing for the German domain. I would expect a short overview of the results from similar studies, which analyzed the trend of heatwave frequency and*

intensity over Germany (i.e. Hundhausen et al., 2023). In this context the added value of this analysis should be presented. It would also be interesting to explore, how the single event of June 2026 has altered the long-term trend statistics. However, this is not an essential addition.

Response: In the revised version of the manuscript, we will report the spatially averaged trends. We will also add the context of previous studies (Tomczyk and Sulikowska, 2018; Hundhausen et al., 2023; Petrovic et al., 2024) and state the added value of our analysis: the focus on early summer and the 1 km grid-point detection.

General comment 4: *For the second part of chapter 3.4 the comparison aims to show, that past heatwaves were similarly extensive or similarly intense to the June 2026 event, but never both. However, this point is entirely undermined by the fact, that the selection criteria are not clearly presented. I have no context for either the 1953, or the 1976 event. Thus, this exemplary comparison does help with putting the June 2026 event into context.*

Response: The two years were selected objectively as the years before 2000 with the largest May–June heatwave duration and cumulative intensity in the Germany-average series: 1976 (12 days, 102.5 °C) and 1953 (11 days, 85.7 °C), compared with 11 days and 127.4 °C in 2026. This criterion will be stated in the text.

General comment 5: *Moreover, it is not clear to me that the selected analogues are a representative sample of the synoptic states from the 25th - 29th of June. I doubt the authors claim that the events have "[…] essentially the same circulation." (l. 338). The quality assurance covers the Z500 residual (Fig. 11 c)), the Z500 pattern distance (Fig. S1) and the sampling uncertainty of the dynamic component of the analogue sample (Fig. 10). From Fig. S1 we can conclude, that the sample (chosen to represent the average state over the 25th – 29th) is significantly more representative of the June 2026 event than the average day on record. I am not able to interpret the absolute values of the Z500 distance. Figure 10 allows for the robust conclusion that the temperature difference between the June 2026 event and the sample is significant. However, neither of these Figures shows that the individual daily samples (N=125) are sufficiently similar of the atmospheric conditions of their respective counterparts. In fact, the Z500 residual between the average Z500 value over 25th -29th of June and their analogue (N=25) is very positive, with values above 75m (Fig. 11 c)). Perhaps, including a sensitivity analysis covering multiple sample sizes and multiple calendar window sizes would be helpful. Combining this with a more detailed analysis of the Z500 residuals could help explaining how much of the T850 anomaly is a thermodynamic effect.*

Response: We agree and we will remove the statement "essentially the same circulation". Distances will be given in physical units: the analogues differ from the observed Z500 fields by 65 m (area-weighted RMS) compared with 115 m for random days, against an event RMS anomaly of 94 m; their median pattern correlation is 0.72, and all lie within the closest 1.5% of candidate days. The sensitivity analysis over sample size (N = 15–40), calendar window (±10 to ±30 days), distance metric, epoch boundary and matching domain will be shown in the Supplementary File. Moreover, the reviewer is right that the positive Z500 residual indicates that the analogues underestimate the amplitude of the ridge: over the target region they reach 62% of the observed Z500 anomaly (131 vs. 211 m).A proper discussion will be added in the revised version of our manuscript.

General comment 6: *Additionally, I see no temporal trend of the T850 anomaly between the 1970-2000 sample and the 2001-2025 sample. Figure 13 c) looks to be around zero on average for both the German and the large European domain. Figure 12 even shows lower values for the 2001-2025 analogues. In stark contrast the authors write: "The late-epoch analogue composite is systematically warmer than the early-epoch composite on all event days […]" (l. 353-354). Additionally, the authors write: "The pooled epoch composites (Figure 13) show that the analogue circulation patterns of the two epochs are nearly*

indistinguishable, while their temperature difference constitutes a coherent warming of the entire domain, except the British Isles." (l. 360-362). In my opinion, they are both equally distinguishable.

Response: We agree; this was an error in the submitted text. The late-minus-early difference is −0.1 °C (95% CI −1.7 to +2.0 °C) and is not significant, also for equal-sized samples, a permutation test and within-epoch analogue selection. In the revised version of the manuscript, we will remove the statements on systematic epoch warming and on the coherent warming of the domain, and describe the Z500 patterns of the two epochs as similar, with small differences in the position and amplitude of the ridge.

General comment 7: *Lastly, the claim of a "[…] new class of intensity-dominated, record-shattering early-summer heatwaves in Germany […]" (l. 24-25) is somewhat disconnected to the rest of the analysis. In line 410-411 the new class is defined as "short" and "record-shattering". Neither "short", nor "intensity-dominated", nor "record-shattering" are defined in the text. In fact, I do not think the word "short" is particularly appropriate. The definition of "record-shattering" from (Fischer et al., 2021) states: "Events that break previous local records by large margins […]". But the record-breaking margins that have been mentioned specifically (i.e., maximum absolute temperature) are in my opinion not large. However, some metric of the June 2026 heatwave may in fact be record-shattering.*

Response: In the revised version of the manuscript, we will remove "new class" and "short/compact" and now define the terms quantitatively. As a quantitative index we will use the mean exceedance per heatwave day (cumulative intensity divided by duration). Following Fischer et al. (2021), we will use "record-shattering" only for the intensity metrics, which broke the previous records by large margins.

Specific comments

C1: *Fig. 1: … the reference period should be mentioned. Fig. 2: … extend the upper bound of Fig. d) – f). Fig. 4: … "Baseline P90" label should be more visible … the red dots showing the 2026 values of the individual regions … I assume "[…] the standard deviation of the regional annual-maximum series […]" is a typo. Fig. 5: The Germany label could be highlighted … I would like to see the graph of Fig. b) and c) for the entirety of Germany in the Supplement. Fig. 8: The units of the color scheme is missing. Fig. 11: Why is the event averaged over 5 days and the analogue composite the average of 25 different … single-day events? …*

Response: Fig. 1 will be redrawn and the the caption will refer to May–June 2026, and the station locations will also be shown. Fig. 2: the anomaly panels will be regenerated with an extended range) and a different colour scale. Fig. 4: will be regenerated with a clearly labelled P90 line. The σ in the Fig. 5 caption indeed refers to the annual-maximum series, which is the correct definition. Thus, the Methods section will be corrected (see our reply to Reviewer 1). Fig. 5a: the Germany label and bar will be highlighted; the Germany-average daily TX for 2026 will also be shown in the Supplementary File. Fig. 8: units will be added.

Fig. 11: analogues will be selected separately for each day, and each observed daily field will be compared with the composite of its own 25 analogues; the event-mean fields in Fig. 11 will be averages of these five daily decompositions, i.e., the observed 5-day mean is compared with the mean of five daily composites, not with a single composite of unrelated days. This is the standard procedure (Jézéquel et al., 2018) and it will be explained in Sect. 2.4 and the caption.

C2: *L. 80-86: … It would be appropriate to mention [other studies] (i.e. Hotz et al., 2024). L. 86-89: Some citation would be helpful. L. 90-92: "Consequently" is used incorrectly. L. 133-134: rationale or citations for the 5-day exceedance. L. 137-141: … I would prefer to see how exactly the choice of reference period affects the results. L. 143-144: The choice of [cumulative intensity] definition is lacking justification (Russo et al., 2014). L. 146: use either "standardized exceedance" or "warming ratio". L. 168-169: ±15-day window, while Jezequel et al. (2018) use ±30 days. L. 216-217: record definition. L. 219-220: how extreme are the*

geopotential-height anomalies? L. 219-222: split sentence. L. 257-258: number of events. L. 401: why "comparatively short"?

Response: All the required changes will be implemented in the revised version of the manuscript and the recommended references will be added.

C3: *L. 425-426: This statement is not backed by Fig. 12 or 13. L. 429-431: … I still do not feel like I have a qualitative understanding whether the 3–4 °C temperature anomaly is primarily driven by the background warming or if it is still the extraordinary circulation amplitude.*

Response: L. 425–426: the statement will be removed (see General comment 6). L. 429–431: in the revised version of the manuscript, we will give a quantitative answer.