Review 1:

In this study, the authors investigate the link between European heatwaves and the large-scale atmospheric circulation through a combination of diagnostics grounded in dynamical systems theory with weather regime analyses. The approach builds on previous work (such as Holmberg et al., 2023) using similar dynamical systems metrics for the investigation of extreme events and extends those by studying seasonal variations and combining the metrics with empirically defined weather regimes that allow for a more accessible interpretation of the large-scale circulation. By doing so, it makes a somewhat incremental, but, in my view, useful contribution to the literature. The results are not clear-cut (e.g., different weather regimes become important during the seasonal cycle), providing evidence of the complexity of the circulation leading to heat waves. The paper is well-written and -structured, and I particularly appreciate the detailed and instructive explanation of the methodology (e.g., the concept behind the dynamical system metrics). Nevertheless, I have a few, mostly minor comments for the authors to consider before I'd recommend the manuscript for publication, as detailed below.

Major comment:

In a recent study, Brunner and Voigt (https://doi.org/10.1038/s41467-024-46349-x) identified pitfalls when studying extreme events through percentiles defined over rolling time windows. If I'm not mistaken, the authors' heatwave definition is based on the approach criticized in this paper. Moreover, Brunner and Voigt emphasize that analyses of seasonal variability, as done here, can be particularly problematic in this context. I thus think that the authors should include a sensitivity analysis showing that the pitfalls identified by Brunner and Voigt do not substantially affect their findings.

We thank the reviewers for their valuable and interesting comments, which improve the quality of the paper.

Our heatwave definition indeed relies on percentiles calculated over 31-day rolling time windows. Brunner and Voigt (2024) examined biases in ERA5 temperature extremes using a 31-day running window and found negative frequency biases mainly in regions and seasons characterized by strong seasonal gradients but weak day-to-day variability. However, they explicitly note that across most of Europe, strong daily variability offsets the seasonal cycle, resulting in weak biases.

According to their Figure S2e (Brunner and Voigt, 2024), the strongest biases in Central Europe occur in March, April, October, and November, while biases from May to September are minor (slightly positive in July, slightly negative in September). Since our heatwave analysis is limited to May–September, we therefore expect any bias to be small and not to materially affect our results.

To prove our expectations, we conducted a sensitivity analysis on the length of the running window by recomputing the heatwaves with the suggested 5d running window (Brunner and Voigt, 2024) for the 90th percentile. In total, 3066 heatwave days were detected using a 5d running window, while 3060 heatwave days were detected using a 31d running window. 3035 heatwave days are common to both definitions. The deviation in heatwave days is thus negligible, with < 1% of heatwave days.

In the revised version of our manuscript, we now refer to this sensitivity study and to the study by Brunner and Voigt (2024) in line 153ff: *Following Brunner and Voigt (2024), who highlighted potential biases from long running-window intervals, we assessed the sensitivity of our results to shorter intervals and found them to remain robust with changes in heatwave days of <1%.*

Minor comments:

- Line 45: Daily mean temperature is more important for the impacts than for understanding the development of heatwaves, right? Maybe mention this here.

We agree that daily mean temperature is more relevant for impacts than for understanding the development of heatwaves and have clarified this in the revised manuscript. L43ff is now phrased

Therefore, soil moisture and surface net solar radiation are key variables to consider in understanding heatwave development, while daily maximum temperature is crucial for assessing impacts.

- L 57: Consider noting already here that the transition between these patterns (advection vs. local processes) in the transition seasons are less-well studied.

We have added a note in this section highlighting that the transitions between these patterns during the transition seasons are less well studied.

L57ff: *The transitions between these patterns during spring and autumn are less well studied. The intricate nature of the relationship between large-scale flow anomalies and temperature and its seasonal variability highlights the need for a generalizable and objective quantification of the strength of this linkage.*

- L 71-74: This is very technical for an introduction section and could be omitted here (it is explained in the methods).

We agree that this might be too technical for an introduction and omitted l 71-74 in the preprint version of the manuscript.

L70-72 in the revised manuscript reads*: It allows for a more detailed investigation of summer heatwaves, as well as the transition months. Though weather regimes provide a useful framework for describing atmospheric circulation patterns, quantifying their link to temperature requires additional diagnostic methods.*

- L 248: I think this summary goes too far. It is not evident form a single example that the metrics and regimes are "clearly connected", as this is only one data point and could likely be a coincidence. I'd suggest using more cautious language here.

We agree that a more cautious language is more adequate here and adjusted it accordingly.

L258ff: *In summary, this case study suggests that inverse persistence and co-recurrence may be linked to weather regimes and, therefore, yield physically meaningful insights, while co-recurrence may also provide valuable information on the relationship between atmospheric circulation and other variables.*

- L 257 and elsewhere: I find it confusing to use "stream" as a short form of "stream function".

We have replaced all instances of "stream" with "stream function" throughout the manuscript to improve clarity.

- Fig. 2: The caption should indicate the fields that are displayed. Furthermore, is there a reason why you always show inverse persistence, although the actual persistence is discussed? This requires the reader to flip everything in their head.

We have updated the caption of Figure 2 to indicate all displayed fields. The second and third sentences of the caption of Figure 2 read now: *The spatial plots (a) and (b) depict the 500hPa stream function and geopotential height anomalies, while (c) and (d) show soil moisture and daily maximum temperature anomalies. (a) and (c) show the mean over 17-21.07.2019 before the heatwave, while (b) and (d) show the peak heatwave mean over 23.7.2019-27.7.2019.*

Regarding the use of inverse persistence, we follow the approach used in previous studies, where the extremal index (i.e., inverse persistence) is typically shown. We have clarified this choice in the manuscript to help the reader interpret the figures correctly.

- L 310: "partly explained" indicates a connection between re-occurrence and persistence, which is not easy to understand without further explanation

We agree that the original formulation could be misleading and have rephrased it for greater clarity. Our intention was to convey that the intuitive expectation—that more persistent atmospheric circulation would lead to similar temperature patterns at the same time and thus higher co-recurrence ratio values—is generally not supported by the data. If such a relationship exists, it appears to be present only in July and August and clearly does not explain the elevated co-recurrence ratio values in the other months. We changed it to:

L.319ff: *In conclusion, heatwaves are characterized by an anomalous high co-recurrence ratio between tasmax and stream function end of May to mid-September, while a higher persistence of atmospheric circulation is only present in July and August. Therefore, a higher persistence of the atmospheric circulation does not lead to more similar temperature and stream function patterns.*

- L 323: "low persistent": I cannot see this in the figure. There is no cluster of points on the right-hand side of the plot.

We understand that the original phrasing might have been misleading and have revised it accordingly. We intended to convey that the highest *tasmax* anomalies mostly do not occur in the lower left corner of the plot, which corresponds to high persistence but a low co-recurrence ratio.

L335f: *In June (Fig. 5d), the strongest tasmax anomalies are either low persistent or highly persistent and with a positive co-recurrence anomaly (like in July and August).*

- L 324: I'd suggest to not mention the weather regimes here, as they are only discussed in more detail later.

We agree that it is better to mention the weather regimes only later, where they are discussed in more detail. We have rephrased the sentence accordingly by omitting ‚*but no tendency for specific weather regime (like in May).*

L335f: *In June (Fig. 5d), the strongest tasmax anomalies are either low persistent or highly persistent and with a positive co-recurrence anomaly (like in July and August).*

- Section 4: I would appreciate a brief discussion about potential linkages of these findings to the mechanisms discussed in the introduction, such as the role of advection throughout the seasonal cycle. This could be added at the end of this section or in section 6.

We agree that linking the findings in Section 4 to the mechanisms discussed in the introduction, such as the role of advection throughout the seasonal cycle, would be valuable. We have therefore added a brief discussion addressing this point.

From a conceptual point of view, one would expect anomalous geopotential (or stream function patterns) and near-surface temperature patterns exhibiting a phase shift of about 90° in winter, whereas summer is characterized by much less of a phase shift (temperature maxima directly below blocking high or only slightly shifted to the west). The level of association is clearly shown to have a distinctive seasonal cycle, suggesting a higher co-recurrence ratio might coincide with less advective situations. However, the co-recurrence ratio does not tell us in which way similar patterns of stream function are linked to similar patterns of temperature, whether there is a distinctive phase shift or not. It can only tell us about the degree of association between large-scale flow and near-surface temperatures, which can either by high in a summer-time blocking (where advection is less important) or in an advection-driven scenario, which is more likely outside of the core summer. Therefore, we cannot fully address the role of advection. One would have to perform a number of additional analyses beyond the scope of our study to combine the dynamical system metric with a measure for advection, analogous to how we combined it with weather regimes in the present work.

We added the following sentence in section 6:

L.531ff: *Further, from the co-recurrence ratio between atmospheric circulation and temperature patterns, the role of advection cannot be directly inferred, as a high association between the two variables may occur both under advection-driven and non-advection-driven conditions.*

- Section 5: The least conclusive part, in my view, is the linkage between maximum and minimum temperature, which is described, but not really interpreted. It would be helpful to add some discussion on potential mechanisms here.

We agree that the linkage between daily maximum and minimum temperature in Section 5 could benefit from further discussion. In the revised manuscript, we have added sentences addressing potential mechanisms, including the roles of radiation, cloud cover, and wind. We also note that a more detailed investigation of these processes could be the subject of future studies. We added the following explanation:

L405ff: *During the core summer months, particularly warm days are expected to coincide with above-average nighttime temperatures due to limited time for nocturnal cooling, and the release of heat trapped during daytime. Consequently, patterns of maximum and minimum temperatures are closely linked, resulting in high co-recurrence values. However, later in the season, in September, day length decreases substantially, and areas experiencing high daytime temperatures may now tend to show average or even below-average nighttime temperatures, thus weakening the link between tasmax and tasmin. Additional factors might further influence nighttime radiative cooling, such as cloud cover and wind.*

- L 517: "might usually represent an atmospheric state close to climatology": I have some doubts about this statement. The fact that the mean over the "no regime" category is similar to climatology does not tell that this is also true for individual cases. In fact, such cases can be quite different from the climatological mean; the reason that they are in this category is that they do not project on any of the patterns of the selected regimes.

We agree that the sentence could be misleading and have revised the wording accordingly.

L537f: *The absence of a weather regime may result from a variety of atmospheric circulation patterns.*

Review 2:

Review of manuscript: "Dynamical System Metrics and Weather Regimes explain the seasonally-varying link between European Heatwaves and the large-scale atmospheric circulation" by Ines Dillerup et al.

Recommendation: Accept after minor revision.

This paper investigates the links between the atmospheric circulation and near-surface air temperature during the extended summer season. The authors use the ERA5 reanalyses and a methodology based on dynamical system metrics and the weather regime approach to study the seasonality of the association between large-scale atmospheric circulation and summer European heatwaves.

The main results of the paper are interesting and deserve to be published. The methods used in the paper are not new but their combination makes the originality of the study. However, I do think that the paper would benefit from a small amount of additional work regarding sensitivity tests and the discussion.

Main comments:

Sensitivity to the choice of the regime approach: the authors have chosen to use the fixed (static) 7-regime approach of Grams et al. 2017. It would be interesting to test a different approach based on a larger number of regimes, for instance with an approach similar as the one (based on sea level pressure) used in Neal et al. (2016) with 30 weather regimes (or weather types). That would allow the presence of more intra-seasonal summer-like regimes (like the Atlantic-low or southerly flow anomaly regime, that is known to be strongly associated with heatwaves over Western Europe, see Vautard et al. 2023 for a recent study) and perhaps of a more refined description of the association between circulation and heatwaves.

We thank the reviewers for their valuable and interesting comments, which improve the quality of the paper.

We acknowledge that the choice of weather regimes may appear somewhat subjective and that our results are limited to the seven-regime framework, compared to the suggested 30-regime approach. For example, the southerly flow anomaly regime you mention is often grouped into the "no regime" category within the chosen seven regimes.

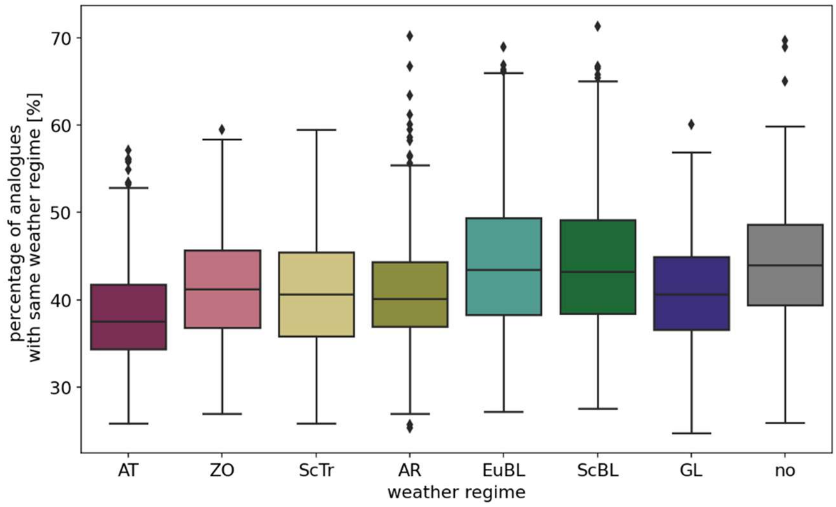
While using a larger number of clusters could provide a more refined representation of intra-seasonal summer-like regimes and their association with heatwaves, this comes at the cost of reduced persistence of individual regimes, which may limit the predictability associated with each regime. The seven-regime definition we adopted has the advantage of being physically meaningful (Hochman et al, 2021), with life cycles that are coherent and interpretable within the framework of dynamical system metrics.

We hope this clarifies our reasoning behind the choice of seven regimes. We agree that exploring alternative numbers of regimes would be interesting, though we feel this goes beyond what can reasonably be addressed in the context of the current revision and might be worth a paper on its own.

If a given regime (its centroid pattern) is strongly connected to heatwaves, I would expect that some of the heatwave properties should depend in some ways of a distance (to be defined, for instance using a simple pattern correlation) between the stream function pattern of the heatwave day and the regime centroid pattern. It would be interesting to see such a diagnostic for the two regime approaches.

The distance between the regime centroid pattern and the atmospheric circulation patterns of every single timestep is represented by the I_{WR} index. We are not certain which specific heatwave properties the reviewer refers to, and we would not necessarily expect these to be related to the I_{WR}. Since heatwaves are extreme events and thus inherently atypical, the most intense events will likely not closely resemble a regime centroid pattern. Moreover, computing a pattern correlation between the regime centroid based on geopotential height and the stream function field would be challenging to interpret, as these represent similar but different physical quantities. For these reasons, we have not performed this additional analysis.

My second comment is about the dynamical system metrics. I know that these metrics have been proposed and used in many recent papers but I have to say that I am still a bit unclear/skeptical on their exact meaning and relevance. With the threshold (2%) you have chosen, you are using 541 analogues per day (this means that you are searching analogues throughout the whole year). It is also well known that given the size of ERA5, even the "best" analogues (based on Euclidean distance or other metrics) are not really good analogues, raising questions about the derived state persistence. I'd be curious to see the mean fraction (relative to the whole set of 541 analogues) of days with the same regime as the reference day. Can you also run the Süveges algorithm on the days of the same regime (as the day of consideration) instead of the analogues obtained with the Euclidean distances? Do you get similar results ?



Please find a plot for the percentage of analogues in the same regime attached (Figure 1) for the summer months, May to September. Note that this is based on the active weather regime life cycle classification and does not represent the I_{WR} index. It is thus possible that the I_{WR} is higher for the individual regimes. Additionally, we examined which regime is most common among the analogues of each reference day and found that it always matches the regime of the reference day. We mention this in the revised manuscript in *l207f: Among those analogues, the most frequent active weather regime life cycle consistently matches that on the reference day, indicating a good quality of the analogues.*

Note that the definition of analogues is here based on the Poisson recurrences. Süveges et al (2007) recommend using 5% as a threshold for recurrences in the computation of persistence, which would include even more analogues per day. Furthermore, the sensitivity of the result to varying thresholds has been tested, yielding a result with little sensitivity to thresholds in a range of 5% to 1% (Faranda et al, 2017).

Selecting only the days that share the same regime as the reference day leads to a varying number of analogues per day. In principle, the Süvegues algorithm could be applied to such a variable sample size; however, this would reduce the interpretability of the results. Using a fixed number of analogues for each day ensures that persistence reflects the average time between clusters of analogues—corresponding to the intuitive understanding of persistence as the duration over which an atmospheric circulation pattern remains similar to itself.

Applying the Süveges algorithm exclusively to days within the same regime, however, would constitute a different research question. In that case, recurrences would be computed for the weather regime centroid, yielding a single Theta value per regime rather than per day—contradicting the concept of instantaneous persistence. Moreover, since Theta represents the extremal index, selecting only days of a given regime would not form a proper time series of extremes and would therefore be conceptually inconsistent with the method's intent. For this reason, we did not pursue this analysis.

My third comment is about the last part of the study and the discussion regarding the soil moisture, surface solar radiation and minimum temperature. I would suggest to reframe a bit this section while trying to provide more explanation behind some of your results. While the soil moisture and surface solar radiation can be considered as additional factors influencing the severity and/or persistence of heat waves, the minimum temperature is part of the heatwave characteristics. I would first describe the results about tasmin, then the ones about the additional factors.

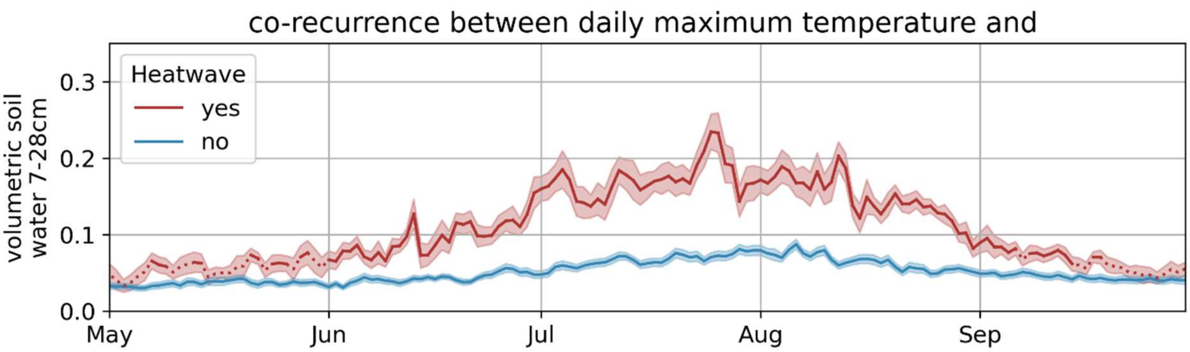
We adopted your suggestion and first describe the results about tasmin, including some mechanisms of how they might interact. We further added some potential mechanisms like cloud cover, radiation, and wind as explanations.

L405ff: *During the core summer months, particularly warm days are expected to coincide with above-average nighttime temperatures due to limited time for nocturnal cooling, and the release of heat trapped during daytime. Consequently, patterns of maximum and minimum temperatures are closely linked, resulting in high co-recurrence values. However, later in the season, in September, day length decreases substantially, and areas experiencing high daytime temperatures may now tend to show average or even below-average nighttime temperatures, thus weakening the link between tasmax and tasmin. Additional factors might further influence nighttime radiative cooling, such as cloud cover and wind.*

Can you explain why there is such large intra-monthly variability during July and August in the co-occurrence of soil moisture with tasmax? Is it just due to the fact that you are only considering the upper soil layer? If you include also the second layer, do you get the same result?

The reason why there is such a large intra-monthly variability during July and August cannot be entirely explained by the use of the upper soil layer only. The co-recurrence between soil moisture in the second layer and daily maximum temperature is depicted in the attached Figure 2. It shows a lower intra-monthly variability during July and August, but a similar pattern in terms of maxima and rising/dropping co-recurrence ratio compared to the soil moisture in the upper soil layer.

Additional factors include the higher variability of the co-recurrence ratio during this time period, visualized in the partly larger standard deviation compared to the co-recurrence ratio of tasmax with surface net solar radiation. This is expected, as heatwave might occur under different soil moisture states, due to prior rain, droughts etc., while the relationship between temperature and surface net solar radiation is not expected to vary as much. We also expect that a quantity with much higher spatial heterogeneity such as soil moisture may also lead to larger variability in metrics such as the co-recurrence ratio. Although soil moisture is thought of as a rather slowly-varying variable, it can change very quickly in the uppermost layers due to a large-scale rain event. We thus suggest a higher number of soil moisture states on different temporal and spatial scales as explanations of the high intra monthly variability. In the revised article, we added a short discussion about the limitations and high variability. L419f: *A large intra-monthly variability of the co-recurrence can be observed, which could be associated with different spatial and temporal scales, as well as spatial heterogeneity.*



When you explain the seasonality and high values of co-occurrence between tasmax and tasmin (line 411), you could also mention the release of the heat trapped during daytime to explain the high values during July and August. Finally, I am not sure than the co-occurrence metric has a direct and simple link with the quantification of the respective contribution of solar radiation and soil moisture to heatwaves (lines 421–422). As the authors mention earlier in the text, no causality and quantification of the effect magnitude can be deduced from the co-occurrence metrics.

We mentioned the release of the heat trapped during the day in the revised manuscript.

L405ff: *During the core summer months, particularly warm days are expected to coincide with above-average nighttime temperatures due to limited time for nocturnal cooling, and the release of heat trapped during daytime.*

We understand that the sentence in lines 421-422 can be misleading and have omitted the affected sentence (L441 in the revised manuscript).

Minor comments:

A general comment: in several instances, a "from" is missing before specific months, see for example line 337: "...but varies only slightly June to August."

We have adjusted the affected instances.

(1) Abstract: page 1, line 9: why spring and autumn here? May and September instead?

This sentence, together with the following one, reflects results for the seasonal cycle of the entire year, as shown in Figure 3. Specifically, Figure 3a shows minima in April and November, and Figure 3b shows maxima in April and November. For this reason, we refer to "spring" and "autumn" rather than to May and September, which would only correspond to the extrema if the analysis were restricted to the extended summer months.

(2) page 3, lines 59: why coupling here? I do not see in the paper any analysis related to the influence of heatwaves or surface temperature patterns on the atmospheric flow. The same comment also applies to other uses of the word coupling throughout the manuscript (e.g page 9, line 246).

We originally used the term "coupling" as a replacement for "co-recurrence ratio" to make the results more intuitive, but we recognize that this term can imply additional relationships that we did not consider. To avoid misinterpretation, we have rephrased the affected sentences.

(3) page 4, line 109: why "throughout the year". throughout the extended summer instead ? Most of the paper concerns the extended summer, and there is very little discussion of the other seasons.

We agree that the main focus of the study concerns the extended summer period. However, the first research question serves as an overarching one for Section 3 (following the case study), which discusses the seasonal cycles of the co-recurrence ratio, inverse persistence, and I_{WR} throughout the year. The subsequent sections address specific research questions that explicitly target the extended summer months. Therefore, we prefer to keep the original formulation of the research questions but have added a sentence clarifying that the study primarily focuses on the extended summer period.

(4) page 5, line 127: please specify how exactly you have detrended the different variables (period, method of detrending etc...)

We have detrended all grid cells separately for daily maximum and daily minimum temperatures using the cdo command detrend, which linearly detrends the time series. We have specified this in the revised version of the paper.

L123ff*: Tasmax and tasmin were detrended linearly to remove long-term trends and seasonality in case of the dynamical system metrics. In practice, we used the Climate Data Operators (CDO) function detrend (Schulzweida, 2023).*

(5) page 5, line 132: "for better comparability" – with what?

The heatwave region used in this study is among those defined by Zschenderlein et al. (2019), who investigated the processes leading to heatwaves in different regions. This choice facilitates a better integration/comparison of our results with the existing literature based on those regions (e.g. Zschenderlein et al, 2019; Holmberg et al, 2023).

L137f: *For better comparability with existing literature, heatwaves are defined in the Central Europe domain from 45°N-55°N and 4°E-16°E (red shading), analogous to Zschenderlein et al. (2019).*

(6) page 5, lines 131–136: I do not see the point of having two different domains for calculating heatwaves and dynamical metrics. It seems to me that using the same domain would be much better. Finally, you have chosen to extend much more the latitudinal than the longitudinal span of the red box to get the dashed box. Can you justify this choice ?

As outlined above, we aimed to ensure comparability with existing temperature analyses and therefore used the same heatwave box as Zschenderlein et al (2019). However, for the dynamical system metrics, the goodness of analogues is crucial. Using the same spatial domain for atmospheric circulation would not adequately capture the large-scale circulation patterns, as these are typically more extensive. This can already be seen in the sensitivity study of the domains (Figure C3), where persistence values are generally higher and only minor differences between heatwave and non-heatwave days are observed. Because the smaller box covers only a limited portion of the atmospheric circulation field, it remains similar to itself for longer periods, even under markedly different large-scale circulation patterns that the smaller domain cannot distinguish.

The choice of the temperature domain follows a similar rationale. Using an area that is too small can lead to questionable analogues, as the similarity of temperature patterns may then be driven by marginal differences in individual grid cells. Moreover, our aim is to link heatwaves to the atmospheric circulation. A larger latitudinal extent is therefore advantageous, as it captures temperature patterns over a broader part of Europe and allows for a clearer distinction between circulation regimes, such as the Scandinavian Blocking (ScBL) and the European Blocking (EuBL).

(7) page 6, line 154: 5% is a rather small area (only a few grid points since you are using 0.5° data) given that your geographical domain is also small.

We chose this threshold based on previous studies, which investigated also the same heatwave region, in particular based on Zschenderlein et al (2019) and Holmberg et al (2023).

(8) Section 2.4, lines 176–178: you haven't clearly mentioned if you classify all days or if you are using a metric with a threshold to identify regime and transition days. Can you detail exactly what you have done? Based on Figure 2e, it seems that you are using the index IWR to define regime and transition days?

We followed the approach of Grams et al (2017). We understand that this part might be confusing due to partly inconsistent terminology that we have adjusted now. By "classification" or "dominant WR type" we mean the previously specified classification of days in their "active weather regime life cycle". We thus classify all days of our time series by following the criteria of active weather regime life cycle in lines 173-176: at least five consecutive days with an I_{WR} greater than 1, within which a local maximum is present that shows a monotonic increase and decrease over the five days preceding and following it. If multiple weather regimes meet these conditions simultaneously, only the regime with the higher I_{WR} is considered active. Whenever those criteria are not met for any of the regimes, the day is classified into the no regime category.

L176ff: *In case that multiple WRs are in an active life cycle at the same time, the one with the highest I_{WR} is chosen as the active weather regime life cycle. It is also possible that no WR life cycle is active when the large-scale flow at a specific time does not meet the criteria. In such cases, its active weather regime life cycle is called no regime. Hereafter, the active weather regime life cycle will also be referred to as weather regime classification, occurrence, or category. Whenever I_{WR} is intended, this is explicitly stated.*

(9) Section 2.5: are you using the full fields or the anomalies for the analogue computation? Have you tested the sensitivity of your approach to other distances (than the Euclidean)?

We have used the full fields for the analogue computation, as this ensures that the analogues are found in similar seasons. Imagine an extreme heatwave in July with positive temperature anomalies of 5K and a warm spell in winter with anomalies of plus 5K, both with a similar spatial pattern of temperature anomalies. The absolute temperatures for the heatwave in July might reach 35°C, while the warm spell in winter might only have 10°C. Those two time points would thus likely be selected as analogues if we applied the methodology to anomalies, which is not desirable in our study. However, we did detrend temperature and used the stream function to remove the climate trend in the timeseries, to avoid a selection bias between the decades.

We have not tested our approach to other distances than the Euclidean, which is suggested by the dynamical system metric approach. For implications of using another distance metric, please refer to Lucarcini et al (2016) and Faranda et al (2019) .

Co-occurrence ratio, lines 209–210: again, coupling implies a two-way interaction and here there is no evidence of such a thing. Furthermore, I do not clearly see why alpha "provides valuable insights into the physical system under investigation".

As outlined above, we have adjusted the wording coupling to avoid misinterpretation. While the co-recurrence ratio cannot provide information about causality, a high co-recurrence implies that similar patterns of different variables often occur together. This in turn implies that there is some sort of underlying dynamics that leads to similar patterns in two variables. Therefore, it does provide insights into the system under investigation. We have adjusted the sentence to provide more clarity.

L219f: *However, since a link suggests the presence of shared underlying dynamics, alpha provides insights into the physical system under investigation.*

(10) page 9, lines 228–229: either "over a 5–day period" or "over 5-day periods"

We have adjusted the sentence to *over 5-day periods* (l235).

(11) page 9, lines 242–247: how do you interpret the strong decline of the co-occurrence starting on the 24, while the heatwave is not yet fully developed?

The co-recurrence ratio reflects how frequently a specific atmospheric circulation pattern and near-surface temperature pattern occur simultaneously throughout the dataset. At the beginning of the heatwave, the I_{WR} of the Scandinavian blocking is the only regime exceeding the threshold value of 1. From the 24th onward, however, the European blocking I_{WR} also exceeds 1, while the Scandinavian I_{WR} rises to exceptionally high values. Although we cannot be certain about this interpretation, we suggest that the subsequent decline may result from a combination of dominant regimes, as well as the very high I_{WR} of the Scandinavian blocking, indicating a less common circulation pattern compared to individual weather regimes. This might lead to a less common co-occurring circulation and temperature pattern. A similar phenomenon can be observed around the second steep decline, near July 28th, when the Greenland blocking gains high I_{WR} values while the Scandinavian blocking I_{WR} remains higher. The very low co-recurrence ratio at the time when the I_{WR} values of the GL and ScBL regimes converge, coinciding with a change in the dominant regime classification, supports this interpretation.

We have added some potential explanations in the revised manuscript l254ff: *We hypothesize that the subsequent decrease in the co-recurrence ratio around the 25th July likely reflects less common atmospheric dynamics with the IWR of EuBL, and later GL, also exceeding 1 and/or the exceptionally high*

IWR values of ScBL. This could explain the reduced frequency with which this atmospheric circulation and its corresponding temperature pattern occur simultaneously throughout the time series.

(12) page 9, lines 248–250: this statement may be true, but at this point in the paper, one case study is not enough to make it true.

We agree that a case study is not enough to make it true and our formulation might be misleading. We adjusted the sentence to improve clarity.

L258ff: *In summary, this case study suggests that inverse persistence and co-recurrence may be linked to weather regimes and, therefore, yield physically meaningful insights, while co-recurrence may also provide valuable information on the relationship between atmospheric circulation and other variables.*

(13) page 9, lines 257–258: "stream" should be "stream function". Note that this need to be corrected in many instances throughout the paper.

We have replaced all instances of "stream" with "stream function" throughout the manuscript to improve clarity.

(14) page 12, lines 301–303: it seems to me that the second sentence and the bootstrap test shown in figure 4b contradicts the first sentence.

We agree that the phrasing of the first sentence might be misleading and have adjusted it in the paper. We meant to express that the only dates, where there is an increased persistence on heatwave days, lie in July and August (and the very beginning of September).

L311ff: *A significantly higher persistence of the atmospheric circulation relative to non-heatwave days is only observed for a few scattered day from the end of June to the beginning of September. The differences between heatwave and non heatwave days are further less pronounced than for the co-recurrence ratio.*

(15) page 17, line 386: change the word "probable" as the role of land-atmosphere interactions and their influence on heatwaves has been shown by multiple studies.

We agree that the role of land-atmosphere interactions and their influence on heatwaves has been shown by multiple studies and have omitted the word probable here.

L397f*: As the anomalous high co-recurrence ratio on heatwave days cannot be attributed to specific weather regimes, we investigate the role of land-atmosphere interactions associated with heatwaves.*

(16) page 18, line 398–399: I would suggest to use a different predictor for the soil moisture availability in order to see whether it can explain the high co-occurrence ratio. What matters is not just the soil moisture anomaly but also how far you are from a shift between soil moisture regimes (dry, transitional, and wet regimes).

We acknowledge that understanding the interaction between soil moisture and near-surface air temperature requires an examination of the underlying soil moisture regimes, which determine the strength of this coupling. The analysis of the co-recurrence ratio between soil moisture and daily maximum temperature can be regarded as an initial step toward assessing the relevance of soil moisture within our current framework. We recognize, however, that this analysis alone is not fully conclusive in identifying the specific states that lead to high co-recurrence ratios between soil moisture and temperature due to the limitations discussed previously. However, we believe that a comprehensive analysis combining co-recurrence ratios with soil moisture regimes would exceed the scope of this review. Nonetheless, such an approach represents a promising direction for future research.

(17) page 18, line 409: "governed". The beginning of the sentence is not really bringing anything new, please rephrase.

We have rephrased this sentence by using the term driven by instead of governed by.

L436ff: *This indicates that heatwaves are driven by different processes and surface net solar radiation is particularly relevant for early summer heatwaves.*

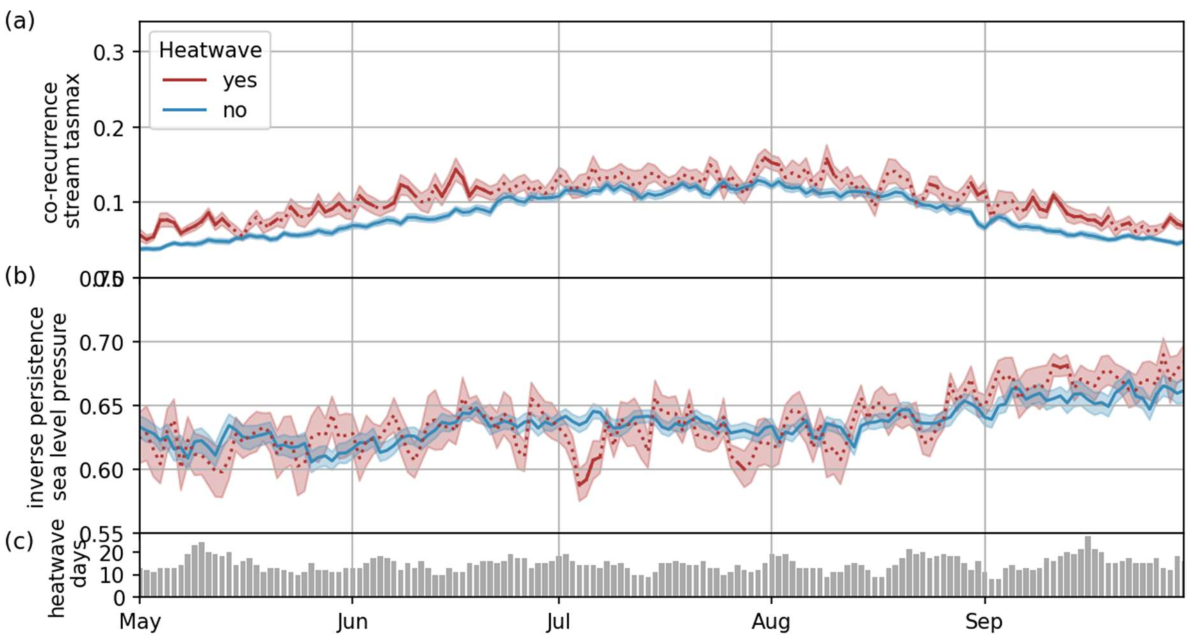
(18) page 18, line 429: I'd rather use "association" instead of "coupling"

We have changed the terminology accordingly. L445ff: *With this aim, we used metrics derived from dynamical system theory, that allow both to objectively identify the persistence of any given atmospheric state and to further obtain an instantaneous measure of the association between two quantities.*

(19) page 20, lines 490-494: it would be very nice to use sea level pressure instead of the stream function to replicate Figure 4.

Please find the replicated Figure 4 using sea level pressure instead of the 500 hPa stream function attached (Figure 3). The Figure shows generally lower co-recurrence values and statistically significant co-recurrence ratios on heatwave days only in May, June, and September. The inverse persistence displays overall higher values (note the different y-axis scale compared to Figure 4 in the draft) and, except for the beginning of July, no statistically significant differences between heatwave and non-heatwave days. These findings are consistent with previous studies on persistence (Holmberg et al, 2023), which reported a weak but significant link between persistent mid-tropospheric circulation patterns and heatwaves in summer, while finding few significant persistence anomalies in surface circulation patterns.

In our view, this result underscores our choice of 500hPa stream function as a viable quantity to describe the large-scale circulation during summer. While surface pressure might more accurately represent smaller-scale details of the near-surface flow field, it is, in our view, not suitable to characterize the large-scale atmospheric flow, in particular during summer heat extremes. Due to strong diabatic heating from the surface, strong positive geopotential anomalies in the mid-troposphere may then co-occur with only little surface pressure anomalies or even the presence of a surface heat low (Fischer et al., 2007). In line with the different roles of atmospheric dynamics, advective processes were also shown to play a much less important role for summer than for winter temperature extremes (Röthlisberger & Papritz, 2023a; Röthlisberger & Papritz, 2023b). Hence, we expect summertime surface temperatures to be less strongly connected to the configuration of the surface pressure field than in other seasons. Using other methods, a number of other studies have also pointed at 500hPa geopotential fields being a much stronger predictors for near-surface temperature extremes than the surface pressure or 1000hPa geopotential fields (Buschow et al., 2024; Suarez-Gutierrez et al., 2020).



(20) page 21, line 517: not always as indicated below and by many previous works looking at the dynamical role of transition days between regimes.

We acknowledge that the sentence *"The absence of a weather regime might usually represent an atmospheric state close to climatology"* could be interpreted in a misleading way, since the agreement between the mean over the "no regime" category and climatology does not imply that this is valid for individual instances. We have adjusted the sentence for clarification.

L537f: *The absence of a weather regime may result from a variety of atmospheric circulation patterns*.

Figures:

Figure 5: middle column: it is very difficult to clearly identify the different colors (specifically the different pink and orange colors). Can you explain why the frequency of GL for all days in August is less than that of EuBL while the climatology of IWR is greater for GL than EuBL?

We selected the current color scheme because the original one proposed by Christian Grams did not meet the journal's requirement of being colorblind-friendly. However, we acknowledge that our chosen colors may not have been optimal. Christian Grams has suggested a revised, colorblind-friendly set of colors for the weather regimes, which we will now implement across all figures. Please find the updated Figure 5 with the adapted colors attached (Figure 4).

The frequency of a weather regime in this context refers to the number of days during which the GL or EuBL regimes exhibit an active life cycle. An active life cycle is defined by the following criteria: at least five consecutive days with an I_{WR} greater than 1, within which a local maximum is present that shows a monotonic increase and decrease over the five days preceding and following it. If multiple weather regimes meet these conditions simultaneously, only the regime with the higher I_{WR} is considered active.

As a consequence of these criteria, it is possible for the climatological mean I_{WR} of GL to exceed that of EuBL, while GL fulfills the active life cycle conditions less frequently or is less often classified as an active life cycle when another regime attains higher I_{WR} values. This situation is illustrated by the July 2019 heatwave (Figure 2e), during which the I_{WR} of GL exceeds 1 around 28 July, yet the active weather regime life cycle classification only switches to GL on 1 August, as ScBL exhibits higher I_{WR} values than GL previously. The life cycle of GL thus starts later and encompasses fewer days than the elevated I_{WR} of GL alone would suggest.