The role of atmospheric circulation changes in Western European warm season heat extremes

Response to reviewer 1:

Review of "The role of atmospheric circulation changes in Western European warm season heat extremes"

The work of Noest et al. uses a similar methodology as Vautard et al. (2023) to investigate dynamical and thermodynamical trends in spring heat extremes in Western Europe. Contrary to Vautard et al. (2023) – who studied summer extremes – they show that in spring most of the trends are thermodynamical. They also focus more specifically on so-called southerly flow patterns and whether those have become more frequent using a flow analogues methodology.

The paper is well written and the subject of research of interest for the community – spring extremes have not been studied a lot in previous works despite the fact, as argued by the authors, that they may be important for soil moisture dynamics and thus for heat extremes in summer. However, the paper is not as strong statistically as Vautard et al. (2023), where the authors used a very large number of climate models and I therefore tend to be slightly sceptical about some of the claims of the authors. Moreover, the general argumentation flow can be improved and thus it is not always clear whether the authors are referring to the analogues of a specific day or to the general total and dynamical trends computed as in Vautard et al. (2023). I recommend strengthening the conclusions obtained before I can advise publication in WCD. I do not know whether it is because of the authors or because of the manuscript format that I received but the visual quality and rendering of the figures can be improved.

We thank the reviewer for reading our manuscript, and are glad that they find the research relevant to the scientific community. We have made revisions and hope the reviewer agrees that our updated manuscript is much improved. Below, we present a thorough point-by-point response to each of the referees' comments as well as a description of the changes that were made in the manuscript.

Regarding the visual quality of the figures, we suspect that this might be caused by the fact that some figures consist of different panels grouped together and used in different versions of Word throughout the entire process of preparing the manuscript. Some figures have already been re-added now, but in case the manuscript is accepted, all figures will be submitted separately as complete files, which we expect to solve this issue.

Major comments:

1. L83: "For both the total and dynamical temperature trends, the trend at a specific gridpoint is considered to be significant when the absolute slope of the regression is larger than 2 times the standard error of the slope." -> I am not sure about the quality of this statistical test. First the authors need to explain how they compute the standard error of the slope. If this standard error is computed correctly then this procedure would be equivalent to a confidence interval around 95% (actually a little bit more restrictive). Why not simply test whether the slope is significantly different from 0 with standard

statistical tests ? Also I want to point out that implementing a false discovery test on these spatially correlated trends would be beneficial, see Wilks (2026).

We have now applied the more standard Wald test to test whether trends are significantly different from 0 on a 95% confidence level. Figures 2, S2, S6 and S7 have been remade with this new approach, with minimal effect on the results. As was requested by the other reviewer, we have also extended Figure 7 with more information on the significance of the trends found by the models, also using this new statistical test. Thanks for the interesting suggestion regarding the false discovery test. To give an indication of the effect of a stricter test, we have applied this to the main ERA5 trends until 2023 (Fig. 2), which is shown in Figure S3. The following changes have been made to the text of the report:

P3,L86: “For both the total and dynamical temperature trends, a Wald test is used to determine whether the trend at a location is significantly different from 0.”

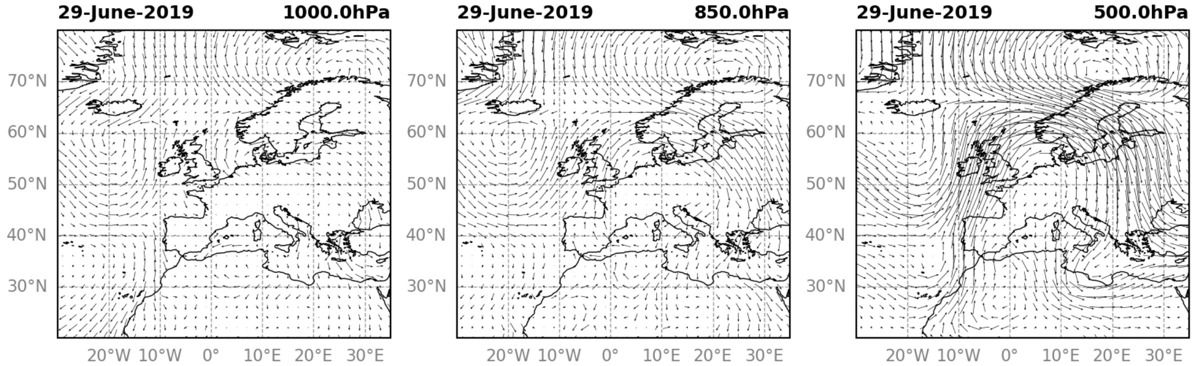
P6,L205: “To give an indication of the effect of using a stricter statistical test, Fig. S3 shows the same results, but applying the Benjamini-Hochberg procedure to control the false discovery rate as described in Wilks (2016). This mainly affects the significance of the dynamical spring trends.”

In SI: “Figure S3: The total trends in the maximum (TXx) and the mean (TXm) of the daily maximum temperatures in a season, for spring and summer, and the dynamical contributions to these trends. Trends are calculated using the ERA5 dataset and are expressed in the amount of warming, in degrees Celsius, per global warming degree (GWD). Dotted areas represent regions where the trend is not significant when controlling the false discovery rate, using $\alpha = 0.1$ since the trends are spatially correlated.”

2. As a general comment, the authors give the 29th of June pattern as the representative of southerly flow patterns: please note that daily streamfunction maps are not suited to evaluate the origin of air parcels because of the non-stationarity of the flow. Also, please note that although the south - south-west origin of air parcels in mid-troposphere is likely given this pattern, on the ground air parcels may not come from the south.

We agree with the reviewer that the ground air parcels do not come from the south (shown in figure below). The following text has been added to the paragraph about the analogue results in the discussion section, to better highlight this limitation:

P15,L355: “Note that, although the persistence of the analogues may provide some indication of the circulation before or after the analogue days, only using isolated daily streamfunction fields in the analogue selection process does not take into account changes in the large-scale circulation leading up to the selected analogue days. Furthermore, the SF pattern is at the mid-troposphere (Fig. 1).”



Wind vectors for the 29-June-2019 at different pressure levels. Data ERA5.

3. Section 2.3: I am not sure what the authors are considering exactly as southerly flow patterns ? Are those analogues of the 29th of June 2019 ? Why considering only this date ? Please explain more how you proceeded exactly.

As is for example done in Thompson et al. (2024) and Faranda et al. (2022), we were looking for a single reference event in order to analyse changes in characteristics of a flow pattern of interest using its circulation analogues. In this case, the analogues of the 29th of June 2019 are thus indeed meant to represent southerly flow days in spring. Therefore, when discussing changes in characteristics, like typicality or persistence, of southerly flow days in spring, this is also based on the analogues. To better clarify this, section 2.3 has been rewritten with more details on how the typicality, persistence and similarity are calculated, rather than only referencing other papers that describe this method, as was also requested by the other reviewer. This hopefully makes it more clear how the analysed characteristics are based on the analogues. The main additions to this section are:

P4,L119: "...The typicality of the event (t_{event}) within a period is defined as the inverse of the sum of the 30 EDs associated with the selected analogues. The typicality of each selected analogue ($t_{analogue}$) is assessed in the same way, by finding its best analogues and using their EDs…"

P4,L125: "...The persistence of the event (p_{event}) is defined as the event day itself, plus the number of consecutive days surrounding the event that have a Pearson correlation coefficient with the event of at least 0.9. The correlation coefficient has been calculated using the streamfunction fields over the same domain in which analogue days are selected. The same is calculated for each selected analogue ($p_{analogue}$)...."

P4,L134"...The similarity for a year is then given by 1 minus the ratio of the selected analogue's ED to the ED of the worst analogue day from both periods combined..."

The 29th of June 2019 was selected as the reference event since this day was part of the hottest June on record for Europe, and its circulation pattern is the most representative of days on which the maximum summer temperature is recorded in central France (Vautard et al., 2020, 2023), as is also mentioned in the report. The latter is also the reason that it is used as the main reference event in Vautard et al. (2023) for analysing changes in frequency and persistence of southerly flow patterns in summer. Since the event is a good example of days on which southerly flow can result in heat extremes, and to make results for spring and summer as comparable as possible, this date was also chosen as the reference event in this study. However, indeed, other days could have been chosen as the reference event as well, which is now added to the discussion section as a potential limitation:

P15,L359: “Finally, instead of 29 June 2019, a different example of an SF event could have been chosen as a reference event, which might have led to different results.”

To introduce the event slightly better, it is now also already mentioned in the introduction section about SF days (instead of only explaining why it was chosen in the method section (P4,L112)):

P2,L44: “This can lead to extreme temperatures, as seen during the heatwaves in June and July 2019, when advection of air from North Africa, caused by subtropical ridges, resulted in temperatures of 46 °C in France and other record-breaking temperatures in the Netherlands, Belgium, and Germany (Sousa et al., 2020; Vautard et al., 2020). **One specific day of the June 2019 heatwave, the SF event on 29 June 2019, will be used later in this study to analyse changes in similar SF days.**”

References:
Faranda, D., Bourdin, S., Ginesta, M., Krouma, M., Noyelle, R., Pons, F., Yiou, P., and Messori, G.: A climate-change attribution retrospective of some impactful weather extremes of 2021, Weather Clim. Dynam., 3, 1311–1340, https://doi.org/10.5194/wcd-3-1311-2022, 2022.

Thompson, V., Coumou, D., Galfi, V. M., Happé, T., Kew, S., Pinto, I., Philip, S., De Vries, H., and Van Der Wiel, K.: Changing dynamics of Western European summertime cut-off lows: A case study of the July 2021 flood event, Atmospheric Sci. Lett., 25, e1260, https://doi.org/10.1002/asl.1260, 2024

Vautard, R., Van Aalst, M., Boucher, O., Drouin, A., Haustein, K., Kreienkamp, F., Van Oldenborgh, G. J., Otto, F. E. L., Ribes, A., Robin, Y., Schneider, M., Soubeyroux, J.-M., Stott, P., Seneviratne, S. I., Vogel, M. M.,and Wehner, M.: Human contribution to the record-breaking June and July 2019 heatwaves in Western Europe, Environ. Res. Lett., 15, 094077, https://doi.org/10.1088/1748-9326/aba3d4, 2020.

Vautard, R., Cattiaux, J., Happé, T., Singh, J., Bonnet, R., Cassou, C., Coumou, D., D’Andrea, F., Faranda, D., Fischer, E., Ribes, A., Sippel, S., and Yiou, P.: Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends, Nat. Commun., 14, 6803, https://doi.org/10.1038/s41467-023-42143-3, 2023.

4. Fig 2: contrary to Vautard et al. (2023) here the authors are considering only two climate models, in this context maybe it would be better to show the results for the two of them separately ?

We have altered the text in the results section and the caption of Fig. 2 to clarify that this figure is calculated using ERA5 data. It now reads:

P6,L179: “Figure 2 shows the total trends in spring and summer temperature extremes, as well as the dynamical contributions to these trends, from the ERA5 dataset.”

P7,L209: “Figure 2: The total trends in the maximum (TXx) and the mean (TXm) of the daily maximum temperatures in a season, for spring and summer, and the dynamical contributions to these trends. Trends are calculated using the ERA5 dataset and are expressed in the amount of warming, in degrees Celsius, per global warming degree (GWD). Dotted areas represent regions where the trend is not significant on a 95% confidence level.”

The model results for the dynamical trends were already shown for each model individually (Fig. S6 & S7), and as requested by the other reviewer, Figure 8 has also been added to the SI for each individual model (Fig. S5).

5. L232: “It becomes clear that the frequency of SF days in spring has slightly increased over time, but that the increase is much bigger for the summer season. Interestingly, the frequencies in July and September have decreased.” -> these results refer to Fig 6. They may be true but I guess that with a limited number of analogues these results may not be significant, please try to test whether this can result from only random sampling. Also, I am not sure I am following what the authors are doing: why are there more analogues in JJA, were you not looking at JJA and MAM independently ?

In all the other analyses in this study, JJA and MAM are indeed considered independently. However, for this specific result, we wanted to investigate how the frequency of southerly flow patterns changes, but also how their occurrences are distributed throughout the warm season. Therefore, we consider all months separately (not MAM versus JJA). To achieve this, a Euclidean distance (ED) threshold is set, and for each month, it is counted how many analogues can be found with an ED below this threshold. The threshold is not set by month but remains a constant value. Thus, we could expect more analogues in JJA when such hot days are more likely (and we expect the change seen, agreeing with Vautard et al. (2023)). To make this more clear, the corresponding method section has been rephrased to:

P4,L138: “As a final method to assess changes in SF patterns, the frequency of SF patterns and how their occurrences are distributed over a longer warm season (MAMJJAS) have been analysed. To achieve this, an ED threshold is used instead of a fixed number of analogues in a season. This time, the entire dataset is divided into a past (1950–1986) and a present (1987–2023) period. For each month in the warm season, the number of analogues with an ED below a certain threshold in that period is counted. The threshold is the same for all months and is taken as the fifth percentile of all EDs in the analysed months of both periods combined.”

To test whether the differences in frequency depicted in Figure 6 are statistically significant, we have calculated the trend in the number of analogues per year over both periods combined for each month, and indeed the trends in spring frequencies are not significant on a 95% confidence level. The trends are visualized in Figure S4, and the following changes have been made to the report:

P5,L144: “To test whether changes in frequency are statistically significant, the trend in the number of analogues per year over both periods combined is calculated for each month. A Wald test is then used to determine whether a trend is significantly different from 0.”

P9,L261: “However, note that June and August are the only analysed months for which the trend in frequency is significantly different from 0 on a 95% confidence level (Fig. S4).”

In SI: “Figure S4: The number of analogues in each year for the different months, with the trend in frequency, the slope of the regression line, and the p-value from a Wald test with the alternative hypothesis that the slope is different from 0.”

6. As a general comment, the authors make big claims for southerly flow in general in their conclusion despite the fact – if I understood correctly – that their results are based on only some analogs of the 29th of June 2019, not all southerly flow patterns that may not be detected by this method. I would therefore recommend to significantly tone down this conclusion. Especially, in L388 the authors claim: “This study shows that Southerly Flow days are becoming more frequent and more intense in spring, contributing to the observed pronounced warming trend” which seems quite contradictory to the results displayed which are much more cautious and do not allow such a strong conclusion.

We have rewritten the sentence, and moved it to a later part of the conclusion as requested by the other reviewer, it now reads:

P17,L431: “The results suggest that Southerly Flow days like the 29 June 2019 event have become more intense and more frequent in spring, although trends in frequency are not yet statistically significant, which might partially explain the observed warming trend.”

To better align with this, the corresponding claim in the abstract has also been rephrased to:

P1,L16: “Our findings suggest that southerly flow patterns, characterized by high pressure over Western Europe and low pressure over the Eastern Atlantic, might become more frequent and intense in spring, which could contribute to the warming trend.”

Minor comments

1. L9 and L13: it would be good to quantify the uncertainties associated to these values

We have calculated the 95% confidence interval for the average Western European trends, and have added them to the abstract:

P1,L8: “Here, the maximum and mean of the daily maximum summer temperatures have warmed 3.3 [2.5–4.2] and 2.4 [1.8–3.1] times faster than global mean temperatures.”

P1,L12: ” Here we show that between 1950 and 2023, the maximum and mean of the daily maximum spring temperatures in Western Europe have intensified 2.2 [1.2–3.2] and 2.0 [1.3–2.6] times faster than global warming respectively.”

The trends were recalculated with confidence intervals as follows:

P3,L87: "The average trends for Western Europe are calculated as the trends in the area-weighted averages of the TXx and TXm values over the land areas within 5° W to 15° E and 45° N to 55° N (Fig. 1, box B). A land mask was derived from the E-OBS dataset (Cornes et al., 2018). A 95% confidence interval for the average trends has been calculated as 1.96 times the standard error of the slope, and is shown in square brackets after the trend estimates."

All confidence intervals are also given in section 3.1 (P6,L179-L207), for example:

P6,L185: "For Western Europe, the average spring TXm trend is 2.0 °C per GWD [1.3–2.6 °C / GWD]."

2. L100: " To ensure the analogue selection is focused on the actual circulation pattern, rather than absolute streamfunction values, all daily streamfunction fields were prepared by subtracting the mean of all streamfunction values within the domain from each individual value within the domain. This preserves the gradients, and thus the wind field pattern as described by the contour lines, whilst removing any differences in absolute values." -> did you remove a spatial mean or the mean removed is different at each grid point ? In the second case the gradients will not be preserved.

The spatial mean is used. This has now been clarified in the text as: "..by subtracting the **spatial** mean..". (P4,L106)

3. For climate model data: were members considered independently to compute the analogues ?

Indeed, members were considered independently. We have clarified this:

P5,L175: "Ensemble members are treated independently in all steps of the repeated analyses."

4. Fig2: total and dynamical trend seem correlated in JJA (which is expected insofar as the latter drives the former), but this is not the case in MAM. Any idea why ?

That is an interesting observation, we believe that during MAM the dynamical component plays a relatively minor role,i.e., other factors such as land-atmosphere coupling, and surface energy balance changes play a stronger influence on temperature variability compared to changes in circulation patterns (e.g. Hoffmann et al. 2021, Ding et al. 2025)

Hoffmann, P., Spekat, A. Identification of possible dynamical drivers for long-term changes in temperature and rainfall patterns over Europe. Theor Appl Climatol 143, 177–191 (2021). https://doi.org/10.1007/s00704-020-03373-3
Ding, X., Chen, G., Wang, Y., & Sun, L. (2025). Demystifying the drivers of the spring warming asymmetry between Eurasia and North America. Science advances, 11(22), eadu2364. https://doi.org/10.1126/sciadv.adu2364

5. Fig8: there is probably a typo in the grammar of the legend. Also, they should be an uncertainty on the ERA5 trend which I do not expect to be small.

The 95% confidence interval of the ERA5 trends have been added to the figure, and the figure caption has been rephrased to make it more clear:

P12,L309: “Figure 8: The distribution of the trends in the maximum (TXx) and the mean (TXm) of the daily maximum spring temperatures, averaged for Western Europe, as found by 105 ensemble members from the HadGEM3 and MIROC6 models. The dashed grey lines represent the trends found in ERA5, with dotted lines showing their 95% confidence interval. ”

References

Vautard, R., Cattiaux, J., Happé, T., Singh, J., Bonnet, R., Cassou, C., ... & Yiou, P. (2023). Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends. *Nature Communications*, *14*(1), 6803.

Wilks, D. (2016). “The stippling shows statistically significant grid points”: How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, *97*(12), 2263-2273.