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

Response to reviewer 2:

Broad context and summary of the key results of the present study:

In Western Europe, the observed trend of summer heat extreme temperatures is found to be up to five times stronger than the global mean warming trend. Earlier research has shown that a considerable fraction of this amplification can be attributed to changes in atmospheric circulation, with a higher frequency and persistence of southerly flow anomalies being a leading cause.

The present paper extends the existing literature by studying heat extremes during the spring months, again aiming to disentangle thermodynamic and dynamical contributions, with an emphasis on the role of southerly flow anomalies. The authors show that the temperatures of the hottest spring day in Western Europe increases on average two times faster than the global mean temperature trend. In line with this trend being substantially lower than the previously reported summer trends, it is shown that the dynamical contribution is on average close to zero in contrast to summer heat extremes. However, the authors find some statistical evidence for a slightly increased frequency and intensity of southerly flow configurations also for spring. Finally, it is reported that some ensemble members of a few selected climate models may capture the observed increase in spring time heat extremes, but on average the model mean underestimates the temperature trends in Western Europe.

General comments:

The research presented in this paper aims to contribute to our understanding of Western European heat extremes by analyzing trends in spring time heat extremes. To my best knowledge, such an analysis that aims to disentangle thermodynamic and dynamical contributions has not been done before for this season. Overall, the paper is well written (with a few exceptions, see further below in the specific comments) and easy to understand. However, there are in my opinion a few issues, mostly with regard to the statistical robustness of the results and the corresponding wording of the implications of the findings. I recommend that the paper has to undergo a major revision before it can be deemed fit for publication.

We thank the reviewer for reading our manuscript and providing such useful feedback. 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 the comments as well as a description of the changes that were made in the manuscript.

Major comments:

Use of a rather limited number of climate models:

In contrast to the work of Vautard (2023), less climate models have been analyzed and therefore the statistical robustness is probably lacking in comparison, particularly given that the results for spring are less inconclusive than for the summer months. An extension of the study to include more CMIP6 models or especially more single model large ensembles is my in opinion too time-consuming, computationally expensive and probably of little additional value. However, the authors should provide some evidence that the models used here perform generally well enough to capture the variability of the atmospheric flow over Europe. One could for instance include a reference to a paper that attests both these models good performances in terms of summer blocking frequency?

The following text has been added to the method section about the models:

P5,L156: "Although global climate models are, for example, known to underestimate the frequency of blocking events, a weather pattern often related to summer heat extremes, the newer CMIP6 models do already generally perform better than older model versions (Davini and D'Andrea, 2020; Schiemann et al., 2020). Palmer et al. (2023) show that the MIROC6 model's performance in terms of summer blocking frequency for Europe is sufficient (satisfactory on a scale with satisfactory, unsatisfactory and inadequate). The HadGEM3 model's performance regarding summer blocking frequency is labelled unsatisfactory, but the model does perform particularly well in capturing the large-scale circulation patterns over Europe for both summer and winter (Palmer et al., 2023). Both models have also been used before in studies investigating European summer temperature trends (e.g., Patterson, 2023) and dynamical contributions to these trends (e.g., Vautard et al., 2023)."

Statistical significance in general:
There are certain instances in the paper (I will specify those below in the specific comment section), where I was not sure if any of the shown differences between past and present era could be deemed statistically significant. Although the results of a multitude of analyses often point toward certain consistent differences between the selected time periods, I still feel that statistical robustness is sometimes not sufficiently proven. In such case, the authors should phrase their findings more carefully and clearly point out a lack of significance in their concluding sections.

As described in more detail in the specific comments sections, we have tried our best to improve the statistical analyses in the report. We have applied a more standard statistical test to the temperature trends, quantified the uncertainty in the average Western European trends, provided more details on the significance of trends simulated by the models, analysed the statistical significance of the changes in frequency of southerly flow days, and toned down conclusions when necessary.

About the flow analogue method in general:
By using just one daily snapshot of the 500hPa stream function, one does not include any information of the evolution of the large-scale flow prior the day of interest. Two very similar looking days in terms of the stream function might therefore feature two very different air masses over the domain of interest. Moreover, concerning the subsequent evolution, one analogue might feature a transient high pressure system whereas the other represents a long-lasting stationary blocking. To some extent, the authors address this problem by also taking account the persistence of the flow field, but this only accounts for what happens subsequently and does not contain information of the days leading up to the SF day. Therefore I would

advise the authors to carefully communicate this potential limitation in their article. Moreover, it should also be mentioned that while the flow during the end of June 2019 heatwave is a great representation of a heatwave-enabling southerly flow anomaly, one could have picked other examples, which would have probably led to different results.

The persistence of an analogue is calculated as the total number of consecutive days with a certain correlation with the analogue, both before and after the analogue day (the complete definition of the persistence has now been added in the method section as well, as requested in your specific comments). In that regard, the persistence could give some insight into the flow before and after the analogue, but this is indeed not taken into account during the selection of analogues.
As is for example done in Faranda et al. (2022) and Thompson et al. (2024), we decided to use a single main reference event, but indeed different dates could have been chosen as well. To better highlight these limitations, the following text has been added to the paragraph about the analogue results in the discussion section:

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). 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."

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

Specific comments:

L30-31:
A few more recent studies with similar inferences could be cited here, such as Rousi (2022) (already mentioned at another section of the manuscript) and/or Dong and Sutton (2025).

Thank you for the suggestions, we have added them as:

P1,L30: "However, Vautard et al. (2023) show that summer heat extremes for Western Europe are warming much faster, up to 5 times faster than the global mean temperature trend. Other studies also find rapid warming trends for European summers and identify Europe as a hotspot for heatwaves (Dong and Sutton, 2025; Rousi et al., 2022)."

L35-36:
I would suggest to add more existing literature about dynamical contributions here such as Singh et al. (2023)

The suggested reference by Singh et al., as well as Rousi et al., (2022) have been added. The sentence now reads:

P1,L36: "However, although these processes are typically less well understood, trends in extremes can also partly be explained by atmospheric dynamical changes (IPCC, 2021; Rousi et al., 2022; Shaw et al., 2024; Singh et al., 2023)."

L86:
Why are you using a land mask derived from the E-OBS data set instead of the one provided by ERA5? Please also clarify how you deal with land-sea masks in terms of the model data with its much coarser resolution.

The decision to use the E-OBS land mask was mainly made because it is also used in the study by Vautard et al. (2023), and we wanted to make the results more comparable when possible. Moreover, the E-OBS data was easily available. Since we are interested in the average trend over land, and in general, the trends over the ocean seem to be lower, only cells that consist entirely of land are considered when calculating the mean with the coarser resolution data. This has now been clarified in the method section about the coarser model data:

P5,L166: "When calculating the mean trend over Western Europe with the regridded data, only cells that consist entirely of land are considered."

L102:
I assume you are referring to a **spatial** mean here? Please clarify.

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

L115 and other:
For better readability, I would suggest to add the definitions for typicality, persistence and similarity here in the paper and not only refer to previous papers.

The section has been rewritten to include all the definitions, with the following being the main additions:

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..."

L178:
The sentence is a bit confusing. I would suggest to rewrite it such as:
In contrast, summer TXm trends are overall lower with 2.4°C per GWD…

Thank you for the suggestion. The sentence has been rewritten to:

P6,L202: "In contrast, summer TXm trends are overall lower, with a Western European average of 2.4 °C per GWD [1.8–3.1 °C / GWD] and an average dynamical contribution of 0.7 °C per GWD [0.3–1.1 °C / GWD] (Fig. 2G and 2H)."

L193:
In my understanding, it is not necessarily always true that an increase in typicality also automatically translates to an increase in frequency. Can you please clarify this?

Indeed, it is not necessarily true that an increase in typicality leads to an increase in frequency, but generally it will be the case. If the most similar analogues are more similar to the event, then there will likely be more analogues above a certain threshold - i.e. an increase in frequency. We have changed the text to phrase this more carefully:

P7,L217: "The typicality of the event in the present period is higher compared to the past period, showing that the closest analogues in the present period are more similar to the event. This likely represents an increase in frequency of similar events – with a likely increase in analogues above a certain threshold, though this is not explicitly shown here."

L231-235:
Please add a statement about whether the depicted difference in figure 6 are statistically significant. I would assume that at least during spring, both periods only show statistically indistinguishable differences in SF day frequency.

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."

L240:
In my opinion, this section needs a short introducing sentence in which you also briefly mention that ERA5 trends were recomputed until only 2014 for better comparison. As it is right now, it is confusing that you start with a comment about ERA5 trends when the section is about the performance of climate models.

To clarify this, the first sentence of the section has been changed to:

P10,L269: "To make ERA5 and model results comparable, the total temperature trends have been recalculated using ERA5 data until 2014 (Fig. 7A and B)."

Figure 7:
There is no indication of whether any of the displayed modeled trends are statistically significant.

The figure has now been extended with panels showing the percentage of ensemble members that find a statistically significant trend for each location. The following changes have been made to the text of the report:

P10,L286: "Whereas both the ERA5 TXx and TXm trends are statistically significant over most parts of Europe, the majority of the modelled TXx trends over European land are not (Fig. 7I). For the TXm trend, most ensemble members find statistically significant trends (Fig. 7J)."

P11,L291: "Figure 7: The ERA5 total temperature trends in the maximum (TXx) and the mean (TXm) of the daily maximum spring temperatures up to 2014, the mean trends found by the 105 model ensemble members from the HadGEM3 and MIROC6 models, the standard deviation of the modelled trends, the percentage of the ensemble members that find a trend as high as the ERA5 trend or higher, **and the percentage of the ensemble members that find a trend that is statistically significant on a 95% confidence level.**"

Figure 8:
You show results here combining the ensemble members from both models into one large sample. I think it would be good to show the results of each model separated or at least provide some additional information in the Supporting information.

Figure S5 has been added to the SI, showing the same results separated for each model. It is referenced to in the report when discussing the differences between the models' results:

P12,L303: "When only taking into account the HadGEM3 model, only one ensemble member finds a TXm trend as high as ERA5, and no ensemble member finds a large enough TXx trend **(Fig. S5)**."

L352:
Just out of personal curiosity: Have you checked whether the extremely strong warming trend in spring has to do with a much earlier retreat of snow cover in this region?

This has not been checked, but it has indeed also been hypothesized by a colleague that the snow-albedo feedback could play an important role in this area.

L388;
Given the rather limited statistical evidence, the claim at the beginning of the conclusion is too bold. Moreover, the conclusion reads quite differently from the abstract, particularly in terms of the ordering of results. I think the found temperature trends should come first, followed by the findings about the dynamical contributions. As the role of changing atmospheric dynamics for spring temperature extremes is overall much less conclusive than for summer, I would also only very carefully state the inferences about possible changes in the southern flow patterns.

The strong claim about southerly flow patterns has been toned down, and has been moved to after the findings about the dynamical contributions. The sentence 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:

Figure 1: Typo in the caption: **Southerly** Flow

Figure 5: Typo in the caption: **Hatched** instead of hashed

L384: for the ".. **for the *model's ability to..*"** instead of "for the models their ability"

Thank-you for spotting these. The typos in both captions have been fixed (P3,L93 & P9,L257). The same typo was made in the caption of Figure 9, this has also been corrected (P14,L344). L384 has been changed to: " .. for the models' ability to.."(now P16,L423).

References:

Rousi, E., Kornhuber, K., Beobide-Arsuaga, G., Luo, F., & Coumou, D. (2022). Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia.

Nature communications, 13(1), 3851.

Dong, B., & Sutton, R. T. (2025). Drivers and mechanisms contributing to excess warming in Europe during recent decades. npj Climate and Atmospheric Science, 8(1), 41.

Singh, J., Sippel, S., & Fischer, E. M. (2023). Circulation dampened heat extremes intensification over the Midwest USA and amplified over Western Europe. Communications Earth & Environment, 4(1), 432.