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.

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).
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.
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.
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 ?
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 ?

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.

Minor comments

1. L9 and L13: it would be good to quantify the uncertainties associated to these values
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.
3. For climate model data: were members considered independently to compute the analogues ?
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 ?
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.

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.