

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

Response to reviewer 1:

I have only minor suggestions for the new version of this manuscript. The authors have increased the statistical strength of their results in a satisfactory way. They are clear that their results do not show a major change in spring patterns (which is not a reason to reject the paper in my opinion), and that the models do not seem to show this pattern. I think the authors should still clarify that this is not clear whether the observed dynamical trends are forced or due to natural variability (abstract and L430 for example). Another small suggestion: in evaluating the trends in the number of analogues per year (Fig. S4) the authors could group years by periods of 10 years to avoid regressing on values between 0 to 3.

We thank the reviewer for going through the manuscript again, and are glad that they find the additions and updates to the manuscript satisfactory. To address the above mentioned suggestions, the following changes have been made to the report: The abstract and conclusion have been expanded:

P1,L20: "Future research could expand on this study by further analysing whether the observed dynamical trends are forced or due to natural variability."

P17,L440: "Note that we do not determine whether the dynamical changes identified are driven by external forcing or internal variability."

Regarding the suggestion about Figure S4, we have now tested grouping the analogues in periods of 10 years. However, this means 4 of the years analysed in Figure 6 are left out of the analysis. Moreover, since in this way you are only left with 7 data points, the 7 decades, the p-value becomes even more sensitive to which 4 years are left out. Whether you leave out the first 4 years, the last 4 years, or shift the 70-year window between those two options decides whether there are 0, 1 or 2 months with a statistically significant trend. To avoid making a decision that might be hard to justify but that does influence the result, we decided to keep the original figure so that at least the time periods analysed are exactly the same for Figure S4 and Figure 6.

Response to reviewer 2:

The authors have addressed carefully all the comments and suggestions of both reviewers. Additional statistical tests were conducted, and a more careful phrasing was used for those findings with only limited statistical robustness. Therefore I would be happy to see the paper published after some very minor revision with regard to the following points:

We thank the reviewer for going through the manuscript again, and are glad that they find the manuscript improved. Below, we present a point-by-point response to the new comments with a description of the changes that were made in the manuscript.

L245:

For the detection of dynamical temperature trends, an additional statistical procedure (BenjaminiHochberg procedure) was applied which accounts for the false discovery rate. The authors only very briefly touch upon this in the text („this mainly affects the dynamical spring trends“). However, the corresponding figure in the Supplement shows that using this test, no grid points show any significant trends any longer. This should be stated more explicitly.

The sentence has been expanded to state the influence more explicitly:

P6,L209: “This mainly affects the significance of the dynamical spring trends, for which all trends become statistically insignificant when applying this stricter test.”

Im am not an expert in those statistical tests, but as far as I understand the applied procedure aims to strike a balance between the often too conservative Bonferroni Correction and a standard uncorrected p-value. In terms of applying it too gridded data, I often wonder whether such an additional test is still too conservative as the spatial coherent pattern of statistically significant grid points generally point to a systematic trend rather than significant trends arising just by chance.

It is true that this procedure leads to a more strict result when applied to spatially correlated trends as opposed to trends that are not spatially correlated. Therefore, we decided to use $\alpha = 0.1$ instead of $\alpha = 0.05$ to partially compensate for the spatial correlation of the trends. In our understanding, this does indeed still result in quite a conservative test. However, just to give an indication of the influence of using a stricter test, we still deemed it useful to include the new test in the supplementary material. Moreover, the influence of picking a higher alpha has also been tested, but, for example, using $\alpha = 0.2$ does not influence the significance of Western European dynamical spring trends. This specific procedure was selected mainly because it is advocated in the paper by Wilks (2016), also for spatially correlated data.

Reference:

Wilks, D. S.: “The Stippling Shows Statistically Significant Grid Points”: How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It, Bull. Am. Meteorol. Soc., 97, 2263–2273, <https://doi.org/10.1175/BAMS-D-15-00267.1>, 2016.

L430:

I think the authors should more explicitly state the implications of the fact that the SF pattern is detected at 500hPa only. In my opinion, that is reasonable choice, but it should be briefly mentioned that air mass characteristics may also depend on the lower-tropospheric circulation.

The discussion section has been edited and extended to include this, and now reads:

P15,L359: “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 patterns are selected based on streamfunction fields at the mid-troposphere level (Fig. 1), without taking into account the lower-tropospheric circulation, meaning the exact origin and characteristics of surface-level air masses could still vary between analogue days.”

L505:

Due to the inserted clause, this sentence could be difficult to understand. I would suggest to rephrase it like this: „The observed warming trend may be partially related to an increased occurrence of

Southerly Flow days like the 29 June 2019 event, although the trend in their frequency is not yet statistically significant.

The sentence has been rephrased using the construction as suggested, only including the

original brief reference to the intensity, like:

P17,L438: “The observed warming trend may be partially related to an increase in frequency and intensity of Southerly Flow days like the 29 June 2019 event, although the trend in their frequency is not yet statistically significant.”

L103: missing space between „to20°“

Thank you for spotting this typo, a space has been added. (P4,L102)