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.

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?

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.

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.

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).

L35-36:
I would suggest to add more existing literature about dynamical contributions here such as 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.

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

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.

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…

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?

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.

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.

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

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.

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?

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.

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"

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.