Manuscript egusphere-2025-4976 *"A combined storyline-statistical approach for conditional extreme event attribution"*

Response to Reviewer 2

The manuscript "A combined storyline-statistical approach for conditional extreme event attribution" by León-FonFay et al. sketches out a new approach for conditional event attribution and demonstrate it with the 2018 European heatwave. They highlight the shortcomings of highly conditioned storylines, and combine them with a less conditioned circulation analogue method. This approach makes it possible to consider forced dynamical trends, and to quantify the changes in probability of occurrence.

I think it's a very important task to address the shortcomings of the individual attribution methods and to produce robust and interpretable attribution metrics that cover a variety of aspects, such as changing risks and intensities simultaneously. This framework provides a new perspective to think about and communicate attribution results, and I recommend it for publication in Weather and Climate Dynamics. While I want to highlight the value of this perspective, I find that the example used in the manuscript shows limitations of the approach which are not sufficiently discussed here. Therefore, I recommend major revisions which address the following points:

Authors' Response:

Dear colleague, thanks for recognizing the value of our study. Your comments have been constructive and helped with the improvement of the manuscript. In the following points, we describe in detail how we have implemented such comments in our study.

Major comment:

1) My main concern is around the interpretation of the results shown in Fig. 6 and 7. When the individual results from the storyline approach and the analogues are brought together, a new pattern emerges which provides the backbone of this new perspective: While the specific 2018 circulation was very extreme compared to the factual analogues (only 1.6% exceeding the event magnitude), the 2018 pattern becomes less extreme compared to the analogues with increasing GWL. We see this in Fig. 6 by the increasing fraction of the tails above the storyline values. Consequentially, the star markers in Fig, 7b don't align vertically as one might expect. This is a bold statement, and I think it should be part of the study to dissect the possible reasons behind this phenomenon, since this structure of the results is what makes up the novelty of the approach.
   Authors' Response:
   Thanks for observing this point. Reviewer 1 has also raised the same concern. In order to address this issue, we have: 1. Explored the role of soil moisture in both approaches, 2. Acknowledge the possible differences between approaches (the atmospheric model used in the storyline approach vs. the coupled model used in

the flow analogues), 3. Rephrased our conclusions to speculations of why we might face an increased probability in warmer worlds, but still stating that our results show that heatwaves with a similar dynamical conditions as the 2018 heatwave become more likely and more extreme due to still unknown mechanisms beyond the fixed circulation, increasing SSTs and GHGs with global warming.

In the manuscript we included this discussion in the new **Section 4.3.1 Increased probability of extreme heatwaves in warmer worlds** and lowered down the tone of our conclusions related to the result obtained through the combined approach:

Lines 5-10

"... Despite no detected changes in terms of atmospheric blocking, the flow-analogue approach further indicates that heatwaves exceeding the storyline-projected intensities become more frequent and extreme at their corresponding warming levels than the factual 2018 event was under present conditions. Specifically, the 2018 heatwave, with an intensity of 2.2 °C and a return period of 1-in-277-years today, becomes a 6.6 °C event with a 1-in-26-year probability in a +4K world. ***This behavior revealed the importance of other physical mechanisms and interactions beyond the atmospheric circulation pattern and thermodynamic conditions influencing the occurrence and intensification of heatwaves.****"*

Lines 351– 386:

" Section 4.3.1 Increased probability of extreme heatwaves in warmer worlds

We hypothesize that this increase in extreme heatwaves could come from a combination of factors: 1) the conservative definition of global warming in the storyline approach (which only imposes changes in SST and GHGs, as they are certain to have a human-induced contribution) restricts the heatwave intensities. 2) the integration of more complex interactions in the MPI-ESM GE (coupled earth system model) in the flow analogues compared to the ones represented by the storylines simulated with ECHAM (atmospheric model with an integrated land component JSBACH). 3) The role of soil moisture as a source of variance in the temperature distributions for future warming levels (Fischer et al., 2012).

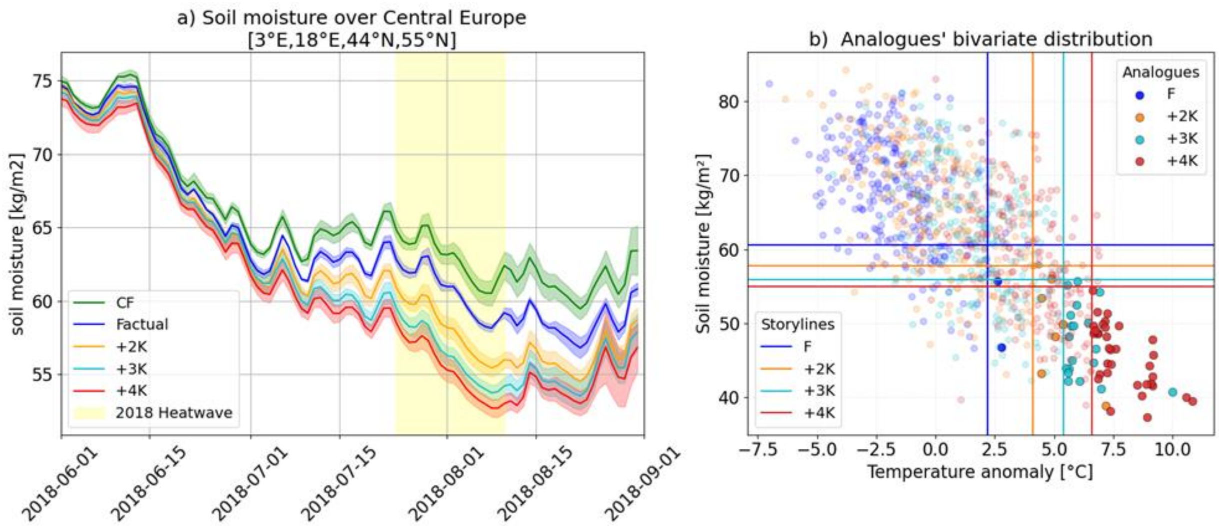

Figure 8. Soil moisture. *a)* Time series of daily soil moisture over Central Europe for each storyline. Solid lines show the ensemble mean, and shaded bands indicate the ±2std range across the five ensemble members for each storyline. *b)* Bivariate distribution of analogue events, displaying soil moisture (y-axis) versus temperature anomaly (x-axis). Colored lines show the soil moisture-temperature values from each storyline. Scatter points represent analogue-derived values, with the darker points indicating analogue events that exceed the soil moisture and temperature thresholds of the corresponding storyline.

The 2018 heatwave was very extreme, not only because of its circulation, but also because of its exceptionally preceding dry conditions (Rousi et al., 2023). In Figure 8, we explore the role of soil moisture in both approaches. In Fig. 8(a), we see how soil moisture is also affected by global warming, decreasing at a faster rate in future levels. The counterfactual storyline had a mean soil moisture of 62.9 kg/m2 during the main heatwave event, while in a +4K storyline, the soil moisture dropped to 54.9 kg/m2. In Figure 8(b), we show the bivariate distribution of soil moisture (SM) and temperature anomaly (T) associated with the flow analogues. The lines in this plot show the corresponding magnitudes in the storyline approach per global warming level. The bivariate comparison demonstrates that the analogues at each warming level occupy temperature(T)-soil moisture(SM) states similar to those in the corresponding storylines and that also their number of extreme heatwaves increases in their own warming levels when soil moisture is included as an additional condition in the analogue attribution (area of bivariate (SM - T) threshold exceedance shown as darker dots in Fig. 8(b)). Here, we can provide evidence that despite having similar atmospheric circulation patterns, the analogues do portray a larger spectrum of events that could emerge due to other underlying conditions, like enhanced soil moisture deficit. It is possible that due to the increasing thermodynamic drying of the future simulations, the role of the atmospheric circulation intensity becomes less important, as more of the analogues also have drought conditions.

Furthermore, the increase of extreme heatwaves in the upper tail of the temperature (Figure 6) and bivariate SM-T (Figure 8) distributions does not come only from the shift in the mean of the distributions towards warmer temperatures due to global warming, but there is also an increase the variance in future levels (see label Figure 6), which could be associated to a strengthened feedback with soil moisture in the analogue catalogue. If one could condition the temperature distributions on the soil moisture of a storyline (i.e. p(SM=60, T)), the variance would be largely reduced. This would lead unfortunately also to a smaller sample size, too small to perform significant attribution.

In sum, both attribution methods are conditioned on the observed dynamics and were made as comparable as possible. The fact that the upper tail of the global warming levels in MPI-ESM increases more prominently does not imply that the approaches disagree. The storyline approach provides a representation of how this single event would change in response to anthropogenic thermodynamic forcing (SSTs and GHGs). The flow analogue approach includes a wider range of heatwaves developing under similar dynamics but different interacting processes, including more extreme heatwaves, probably resulting from a reduced soil moisture availability. Even though these results may be model-dependent, it is worth emphasizing that both approaches rely on the same atmospheric model. The storylines are simulated with ECHAM6, the atmospheric component of the MPI-ESM model, while the flow-analogues are extracted from scenario simulations of MPI-ESM. Including additional models for comparison in future work could provide more detailed insights on this matter. "

Lines 408 –415:

"Warmer storylines of the 2018 heatwave are not as rare at their corresponding level of global warming, as the 2018 heatwave was in the present. While the heatwave intensities increase linearly per degree of global warming, their frequency exponentially increases. Specifically, the observed 2018 heatwave had a 2.2 °C intensity with a return period of 1-in-277 years in the present time, intensifying to 6.6 °C with a return period of 1-in-26 years in a +4K world. ***The reason for the increased intensification in future climates is subject to further studies. So far, we assume that these differences could emerge from the conservative representation of global warming in the storyline approach, soil-feedback mechanisms enhanced in the MPI-ESM grand ensemble that leads to a larger variance with global warming, and other processes accounted in the coupled model"***

Lines 459 – 462:

"The major outcome of this study is the combination of two conditional attribution methods, namely storylines and analogues. In this way we can provide information how a specific historic extreme event would have occurred without anthropogenic warming and how it

might behave for future climate states in combination with statistics like probabilities of occurrence for classes of similar extremes..."

Lines 467 – 473:

*"**In Section 4.3.1 we discuss the possible reasons for this increase in probability.** According to Feser et al. (2024), the spectrally nudged storyline approach is conservative as it represents the influence of anthropogenic global warming solely through information stored in sea surface temperature and greenhouse gases, while other variables, which are likely also influenced by human activity -like aerosols- are not taken into account, representing climate change more cautiously. **This may restrict heatwave intensity, which accompanied by the role of soil moisture in temperature variance and other unknown processes coming from the coupled model may explain the increase in the upper tail distribution of the heatwaves detected in the analogue approach, compared to the simulated storylines."***

Lines 483-489:

***"Naturally, it is not only the atmospheric circulation pattern and global warming that define the characteristics of a heatwave.** For such reason, not all flow-analogues evolve into a heatwave, **and others get much more intense than the observed one.** Many other factors and interactions play a role, **like preceding drought conditions.** Moreover, by combining the storyline and flow-analogue approaches, we are not only able to project heatwaves under future warmer scenarios but also to assess their likelihood. **This synthesis allows us to state that future events with a similar blocking system to the 2018 heatwave might become more extreme and more frequent in future climates than the 2018 event was in the present**, highlighting a critical shift in environmental risks as global temperatures rise."*

a) I can imagine possible physical explanations for this phenomenon: 2018 stood out not just by the circulation during the HW but also by the preceding precipitation anomaly. This led to dryer soils during the HW which exacerbated the impacts of the 2018 circulation. Due to the increasing thermodynamic drying of the future simulations, the role of the circulation intensity becomes less important, as more and more of the analogues also have drought conditions. You could test this by comparing the temperature anomaly of the 2018 circulation to the other summer temperature anomalies within each storyline scenario. If there is such an underlying physical explanation, the temperature anomalies of the 2018 circulation would move closer to the mean with an increasing GWL.

Authors' Response: Thanks for your interpretation of the possible underlying mechanisms behind the increased probability of the 2018 heatwave in warmer worlds. Both in our discussion with Reviewer 1 and in the manuscript, we have analyzed the role of soil moisture in both approaches. Your argument aligns with the finding in our new Fig. 8: analogues show soil conditions even drier than the corresponding storylines. Specifically, we like your interpretation, which has been included in the manuscript.

Lines 369-371:

"It is possible that due to the increasing thermodynamic drying of the future simulations, the role of the atmospheric circulation intensity becomes less important, as more of the analogues also have drought conditions."

In the previous point, we addressed the discussion on this point as a possible reason for the increase in extreme heatwaves, together with Reviewer 1's input highlighting that soil moisture deficit leads to an increase in variance of the analogue's distributions.

Lines 467 – 473:

*"**In Section 4.3.1 we discuss the possible reasons for this increase in probability.** According to Feser et al. (2024), the spectrally nudged storyline approach is conservative as it represents the influence of anthropogenic global warming solely through information stored in sea surface temperature and greenhouse gases, while other variables, which are likely also influenced by human activity -like aerosols- are not taken into account, representing climate change more cautiously. **This may restrict heatwave intensity, which accompanied by the role of soil moisture in temperature variance and other unknown processes coming from the coupled model may explain the increase in the upper tail distribution of the heatwaves detected in the analogue approach, compared to the simulated storylines."***

Lines 483-489:

*"**Naturally, it is not only the atmospheric circulation pattern and global warming that define the characteristics of a heatwave.** For such reason, not all flow-analogues evolve into a heatwave, **and others get much more intense than the observed one.** Many other factors and interactions play a role, **like***

preceding drought conditions. *Moreover, by combining the storyline and flow-analogue approaches, we are not only able to project heatwaves under future warmer scenarios but also to assess their likelihood.* ***This synthesis allows us to state that future events with a similar blocking system to the 2018 heatwave might become more extreme and more frequent in future climates than the 2018 event was in the present****, highlighting a critical shift in environmental risks as global temperatures rise."*

b) However, I think there might be a different explanation for the observed pattern, as I don't agree with the statement that the "approaches are physically consistent, as they rely on the same model" (L332). While the models share the atmospheric component, they differ in the ocean (fully coupled, vs AMIP based on observations scaled to warming patterns from large ensemble), and more importantly in the atmospheric dynamics (ECHAM6 vs nudged-to-NCEP). Small differences in climatology or warming rate between storylines and fully coupled grand ensemble would translate to a shift in the slope of the stars in Fig. 7b. I think this is important to mention (or you could compare the ECHAM_SN climatology from (Schubert-Frisius et al., 2017) to the climatology in the grand ensemble?).

Authors' Response:

We agree with the reviewer regarding the descriptions of the differences between the model set-ups. We could unfortunately not couple the ocean, as, not being nudged due to a lack of high-quality observation data to nudge to, it would deteriorate and constantly work against the atmosphere which is nudged (and we did not want to have to frequently reinitialize the model because of that). The existing difference in atmospheric dynamics between storylines and the grand ensemble should not have much effect in this study as we compare only similar weather patterns. Both attribution methods are conditional on the observed dynamics and were made as comparable and similar as possible, that's what we meant with 'physically consistent'. We included in the text the use of climatology of Schubert-Frisius et al. (2017) in this study. We compared this climatology against the one of the grand ensemble (Fig. R1), showing similar magnitudes over the summer months used in this study (June-August). Finally, we think that the differences between climatologies should not be relevant, since we are comparing anomalies between equivalent datasets. For each ensemble member we calculate the temperature anomalies between future climate and the corresponding climatology of the same member.

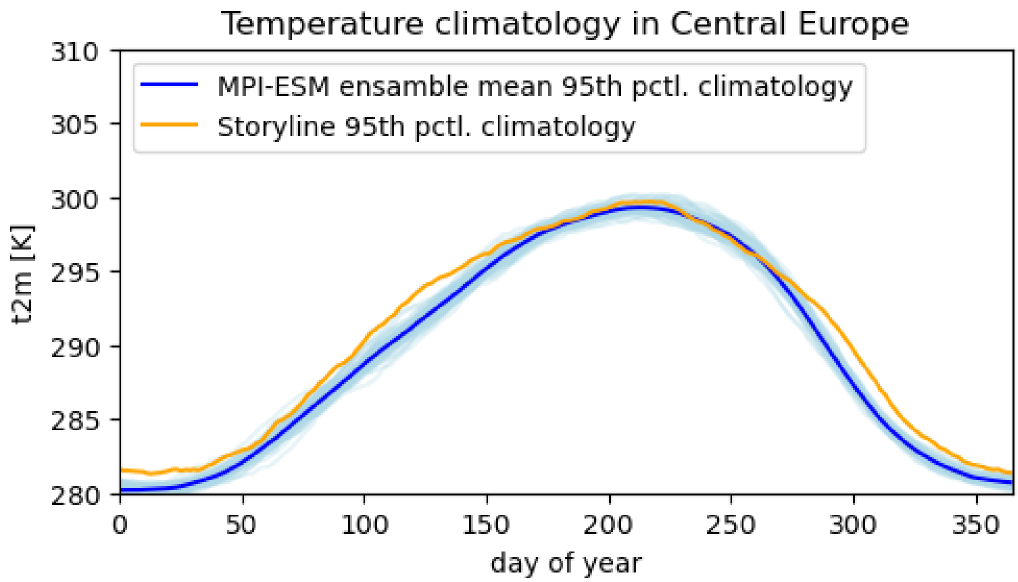

Fig. R1. Climatology comparison between storylines and MPI-ESM over Central Europe. Here we show the 95th percentile of the climatology (1985-2014) used to define a heatwave. Thin light blue lines correspond to each ensemble member, and the dark blue lines correspond to the ensemble mean. The orange lines correspond to the 95th percentile of the same climatological period spectrally nudged to NCEP reanalysis data.

Lines 138-140:

"We use the long-term ECHAM_SN simulation also spectrally nudged towards NCEP reanalysis data (Schubert-Frisius et al., 2017) as a climatological period (1985-2014) consistent with our simulations."

2) Selection of the analogs: From how I understand your approach, the main idea is to loosen the conditioning around a storyline in order to consider 'similar' events as well. However, it is not trivial to me to what extent the definition of similar will influence the outcome of this analysis. My impression from (Rousi et al., 2023) is, that the dynamical conditions that caused the 2018 HW were quite exceptional. It's understandable that these conditions must be loosened, but while your framework relies heavily on the concept of 'similar' circulation analogues, you set the definition in a way that defines the 2018 circulation as a 1-in-4 years event. Resulting from this, 83.5% of the analogues don't produce a HW under the factual climate. I'm not arguing that these conditions need to be stricter, but I would like to see a discussion on how the strictness in the conditioning of the analogues impacts a) the quantitative results, and b) how much the interpretation of the results hinges on the concept of 'similarity'.

I also agree with point 2b by reviewer #1, and think that a conditioning on the temporal pattern would pull away some weight from finding analogues with the right intensity (As seen in Fig. S4 the analogues are a bit on the weak side. For me, it would have helped to see some examples of the analogues and their related temperature patterns).

Authors' Response: We agree that the definition of "similar" atmospheric circulation is central to our framework. As also discussed in response to Reviewer 1 (point 2b), we designed the analogue detection algorithm following advice from our colleague (and flow-analogue method developer (Zorita and von Storch, 1999)) Eduardo Zorita. Our analogue-selection method intentionally balances two competing needs: retaining the temporal evolution of the 2018 pattern, and keeping a sufficiently large pool of good analogues. As also argued in our previous response to Reviewer 1, we can apply this 2-step filter due to the large sample we have available (92000 days: 92 summer days x 20 years x 50 members). Still, we did a brief analysis to check what if we looked for analogues based on their 5-day mean pattern instead of their sequence. In Fig. R2 we show that the temperature anomaly distributions are unaffected in shape, but using a 5-day mean provides a larger sample. It is very likely that the analogues identified by the 5-days sequence threshold are a subset of the ones identified by the 5-day mean threshold. In a way, all analogues identified through a sequence contain temporal evolution and high mean correlation, while the ones identified by the mean pattern contain high correlation and larger sample while including patterns evolving differently to the one of interest. Therefore, given that we obtain a large enough sample and more detailed information using our initial method, we find our more adequate for the aim of our study.

Evidently, quantitative results (ex: number of analogues) are sensitive to the thresholds and definitions, but we focus on the relative changes between global warming levels.

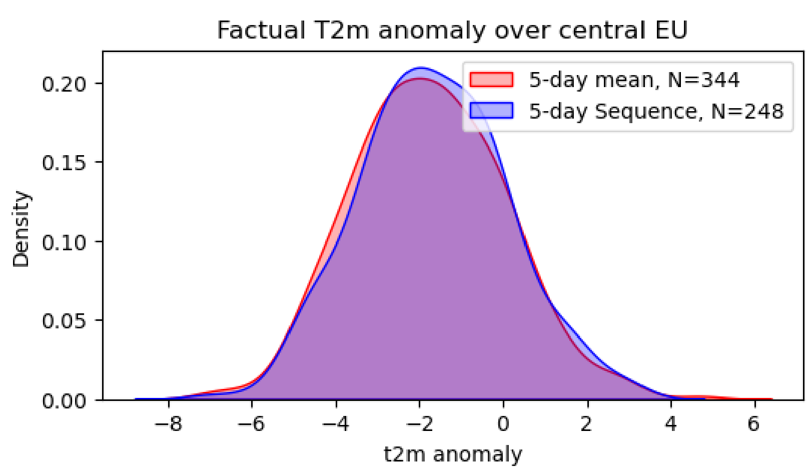

Fig. R2 Analogue detection conditioning. The blue distribution corresponds to the analogues identified in the MPI-ESM GE as described in the manuscript using a 5-days sequence as a reference pattern. The red distribution corresponds to the analogues identified using a 5-day mean of the reference pattern. N shows the number of analogues.

3) Forced circulation changes: The approach relies on a correct representation of P(D) by the grand ensemble. The fact that this is a topic of high uncertainty is one of the main reasons for using the storyline approach. While (Vautard et al., 2023) don't go into a separation of forced and unforced dynamical trends, their results suggest, that no member of the MPI-ESM GE is able to reproduce the trends in European circulation patterns related to summer heatwaves. Since the advantage of the proposed framework is its consideration of circulation changes, **I think it should be discussed what we can realistically expect from CMIP6 models (Shaw et al., 2024).**

   Authors' Response:

   We have included a brief discussion in the manuscript.

   Lines 439 – 446:

   "Recent studies have shown that all nearly all CMIP6 model simulations fail to fully capture the accelerated trend in western European heat extremes, mostly by underestimating or totally missing the circulation-induced contribution related to more frequent southerly flow anomalies (Kornhuber et. al, 2024, Singh et al., 2023, Vautard et al. 2023). It has to be kept in mind that our combined attribution is designed to assess the impact of climate change on individual observed extreme events. While we account for possible dynamical changes in the future, we only do so by estimating the future probability of an event that is dynamically similar to the observed one. Our method therefore cannot provide more general inferences about expected changes in the dynamical contribution – either forced or from internal variability – that may dampen or amplify future heat extremes."

To summarize: While there are currently limitations to the feasibility of the proposed approach, which are mostly based on the physical consistency between the two elements, I don't think that these constraints devalue the usability of this framework.

Authors' Response:

Thank you for the positive evaluation of our framework. We now tried to better explain and discuss the limitations of the feasibility of the method.

Minor comments:

1) It looks like the model output and ERA5 have a 12h offset (while they both refer to the daily means). You could correct this to make the plot look nicer.

   Authors' Response:

   Thanks for your observation, we've corrected Figure 2.

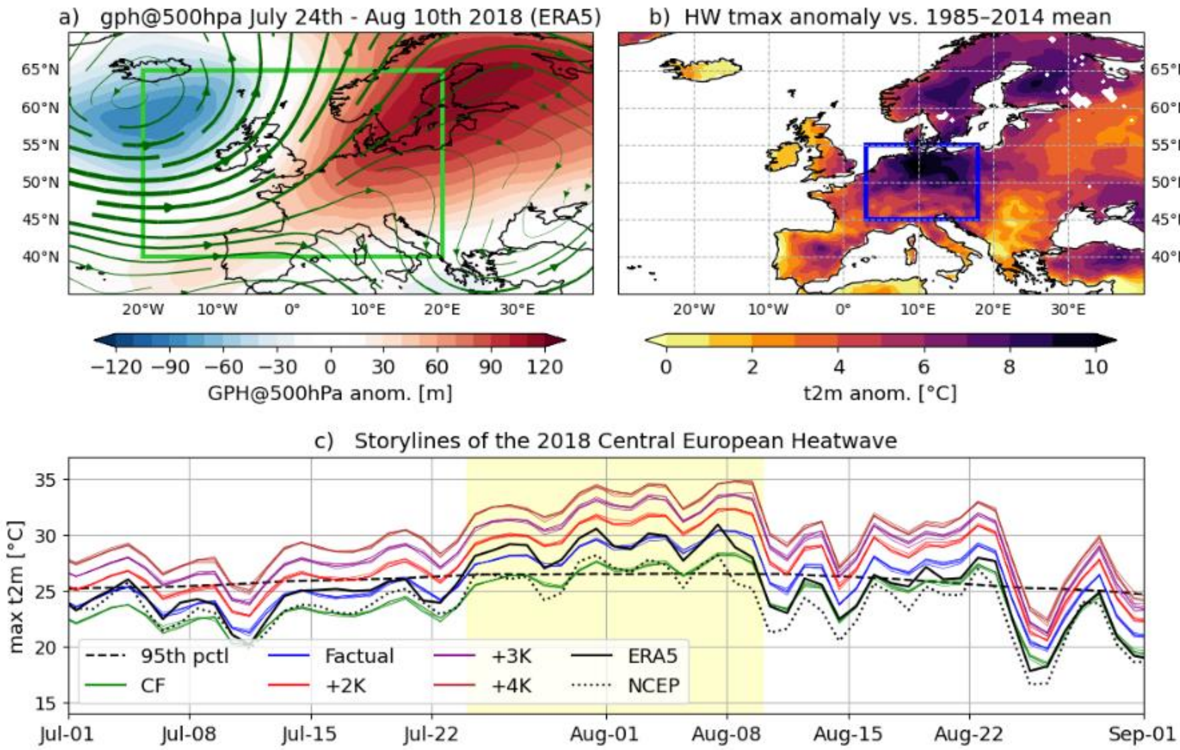

Figure 2. The 2018 Central European heatwave...

2) The reference by (van Garderen et al., 2021) has the wrong year.
   Authors' Response:
   Thank you, we have corrected the reference.

3) The event definition (blue box in Fig. 2) is based on the impacts of the 2018 circulation, but the key point is to loosen the circulation restrictions, thereby including circulation analogues with different spatial temperature patterns. So I wonder if it would make sense to use a bigger box for the second part of the analysis?
   Authors' Response:
   We consider that keeping the same region of study is crucial for the consistency between methods, otherwise we would not be attributing changes in the same region. Besides, in the same box (Central Europe), different spatial temperature patterns can emerge. This is evidenced by the obtained temperature distributions, where many different magnitudes are observed, coming from a wide range of possible spatial patterns occurring under similar dynamics.

Best regards,

Istvan Dunkl

References:

van Garderen, L., Feser, F., and Shepherd, T. G. (2021). A methodology for attributing the role of climate change in extreme events: a global spectrally nudged storyline, Natural Hazards and Earth System Sciences, 21, 171–186, https://doi.org/10.5194/nhess-21-171-2021.

Otto, F. E. (2017). Attribution of weather and climate events. *Annual Review of Environment and Resources*, *42*, 627–646. https://doi.org/10.1146/annurev-environ-112621-083538

Schubert-Frisius, M., Feser, F., Storch, H. von, and Rast, S (2017). Optimal Spectral Nudging for Global Dynamic Downscaling, Monthly Weather Review, 145, 909–927, https://doi.org/10.1175/MWR-D-16-0036.1.

Shaw, T. A., Arblaster, J. M., Birner, T., Butler, A. H., Domeisen, D. I. V., Garfinkel, C. I., Garny, H., Grise, K. M., and Karpechko, A. Yu. (2024). Emerging Climate Change Signals in Atmospheric Circulation, AGU Advances, 5, e2024AV001297, https://doi.org/10.1029/2024AV001297.

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 (2023).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.

Rousi, E., Fink, A. H., Andersen, L. S., Becker, F. N., Beobide-Arsuaga, G., Breil, M., Cozzi, G., Heinke, J., Jach, L., Niermann, D., et al. (2023). The extremely hot and dry 2018 summer in central and northern Europe from a multi-faceted weather and climate perspective, *Natural Hazards and Earth System Sciences, 23,* 1699–1718. https://doi.org/10.5194/nhess-23-1699-2023

Zorita, E., & von Storch, H. (1999). The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. *Journal of Climate, 12(8)*, 2474-2489. https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2.

Kornhuber, K., Bartusek, S., Seager, R., Schellnhuber, H. J., & Ting, M. (2024). Global emergence of regional heatwave hotspots outpaces climate model simulations. Proceedings of the National Academy of Sciences, 121(49), e2411258121. https://doi.org/10.1073/pnas.2411258121

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. https://doi.org/10.1038/s43247-023-01096-7