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:

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.
   - 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.
   - 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?).

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).
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).

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.

Minor comments:

- 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.
- The reference by (van Garderen et al., 2021) has the wrong year.
- 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?

Best regards,
Istvan Dunkl

References:

van Garderen, L., Feser, F., and Shepherd, T. G.: 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, 2021.

Rousi, E., Fink, A. H., Andersen, L. S., Becker, F. N., Beobide-Arsuaga, G., Breil, M., Cozzi, G., Heinke, J., Jach, L., Niermann, D., Petrovic, D., Richling, A., Riebold, J., Steidl, S., Suarez-Gutierrez, L., Tradowsky, J. S., Coumou, D., Düsterhus, A., Ellsäßer, F., Fragkoulidis, G., Gliksman, D., Handorf, D., Haustein, K., Kornhuber, K., Kunstmann, H., Pinto, J. G., Warrach-Sagi, K., and Xoplaki, E.: 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, 2023.

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

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.: Emerging Climate Change Signals in Atmospheric Circulation, AGU Advances, 5, e2024AV001297, https://doi.org/10.1029/2024AV001297, 2024.

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.: 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, 2023.