Reviewer 2

This paper presents a methodology for attributing extreme weather events using flow analogues, using three different event types within Europe (precipitation, heat, and wind) as worked examples. The method is compared to the 'traditional' probabilistic attribution method. The paper provides arguments for the methodological choices in the analogues attribution technique, stressing both the advantages compared to GEV and advantages compared to earlier uses of analogues. The approach methodology presented is well, and is a great development for the use of analogues in attribution.

However, I have a few major comments that I think need addressing, followed by some more minor comments.

We thank the reviewer for their comments. In the following we address them.

1. I think more consideration of the results and how/why they initially appear to (in some cases) contradict the GEV method needs to be made. More thought and discussion on the 'attribution statement', and importantly the framing of the questions that each attribution methods can be used for, needs to be made. A related point which is not discussed in detail in the paper, is the advantage of the method presented over other conditional methods (e.g. those referenced in line 48), and how using a range of methods to provide multiple lines of evidence could be useful. As is, the paper appears to pit the analogues method against GEV, rather than consider how it is useful to consider both to increase understanding.

This could be a useful reference - Coumou, D., Arias, P.A., Bastos, A., Gonzales, C.K.G., Hegerl, G.C., Hope, P., Jack, C., Otto, F., Saeed, F., Serdeczny, O. and Shepherd, T.G., 2024. How can event attribution science underpin financial decisions on Loss and Damage?. *PNAS nexus*, *3*(8), p.pgae277.

We thank the reviewer for suggesting this interesting reference which is indeed very relevant for our discussion. We have included at multiple places in the paper comments to answer the questions raised by the reviewer:

a. Framing of the questions:

i. Introduction L43-60: "This method is unconditional in the sense that it is purely statistical and gives absolute probabilities for the yearly maxima of the observable X of interest. Whatever the actual dynamics of the observed event, it compares its intensity with the yearly maximum intensities of the past. It therefore bears the risk of comparing events that were yearly maxima but that had different dynamical mechanisms. To alleviate this issue, several conditional methods have been proposed [...] they address the question of the mean changes between the counterfactual and factual worlds for events dynamically similar to the one observed:[...] In this sense, these methods condition on the dynamics to isolate the thermodynamical signal. Conditional methods are also useful to explore the physical causes of changes in the extremes, one of the key elements to support the results of attribution studies.

This conditional attribution framework allows to answer the question: how a similar large-scale circulation pattern in the two worlds leads to different outcomes in an observable of interest? [...]"

ii. Discussion L 359-374: "We nevertheless want to note that the main drawbacks of the method presented here are also common to the classical EVT method, namely: under-sampling, representation of natural variability, use of past observations vs model outputs. The interpretation of the results of the two approaches are also different. The EVT-based approach gives the probability that the yearly maximum of an observable is above a given level and therefore the probability ratio gives how this probability has changed between the factual and counterfactual worlds. It thus encompasses both the dynamical changes --- increasing frequency of certain weather patterns caused by anthropogenic global warming \citep{vautard2023heat, faranda2023atmospheric, dong2024indo, d2024summer} --- and the thermodynamical changes for the strongest extremes. As a consequence, its analysis may be far from the actual extreme event observed and in particular can always make an attribution statement even though the dynamics of the event has never been observed in the past at the place considered \citep{faranda2023climameter}. The flow analogues method on the other hand gives the change in the probability of a certain level given the synoptic pattern, as soon as the synoptic pattern has good analogues in the past. This method separates the dynamical contribution from the thermodynamical contribution. It does not address the unconditional probability of reaching an extreme --- which may be the most interesting aspect for the general public --- but it tends to give a better attribution signal because thermodynamical changes are likely more easily detectable than dynamical changes \citep{shepherd2014atmospheric, vautard2023heat}. Our results suggest that the EVT-based approach may tend to be too conservative in its attribution statements by considering only the strongest extremes for which rare or very rare dynamical mechanisms may overrun the climate change thermodynamical signal."

b. Advantage of our method: L388: "However, conditional attribution methods, such as the one presented here or others \citep{yiou2017statistical, terray2021dynamical, de2024western, leach2024heatwave}, are more focused on the very dynamics of the event observed and may provide a more detectable (thermodynamical) signal. The main advantage of our method is to use only past data and therefore to avoid common pitfalls of modeling studies."

c. Multiple lines of evidence: L390: "As argued recently by \citet{coumou2024can}, using a range of methods to provide multiple lines of evidence for an attribution statement useful for practitioners is therefore absolutely necessary."

2. For the precipitation event many of the analogues show little rainfall (particularly after detrending). This suggests that the analogues poorly represent the precipitation of the observed event. Can you show that the analogues do represent the precipitation adequately? (This could also be needed for temperature and wind, but it appears less of an issue for these variables – perhaps suggesting the method is only suitable for specified event types).

We thank the reviewer for this suggestion and we have now included it in the paper (Figure 1). We have also changed the region where the analogues of the precipitation event were looked for as the resulting precipitation field did not reflect the event itself.

We also refer the reviewer to our answer to reviewer 1 for the particular case of the precipitation event on the points raised.

3. One point that is mentioned in the discussion (line 363), but I feel should be stressed further, is that the method only works when there are good analogues. Further discussion of this would be valuable – are there certain event types better suited to the method? Or some events (perhaps hurricanes) where the event should not be used due to insufficient past analogues (though maybe large model ensemble could be used instead)? What if good analogues only occur in later decades?

The reviewer raises interesting points that may need further investigation. We have discussed partly some of the points raised, but without a more systematic investigation we prefer not to give general conclusions which may reveal not to be true. We feel like the quality metric we provided and the discussion on the quality of the composite fields (L197-203) for the analogues provides evidence for how to determine whether good analogues are found for a particular case study. To answer more precisely the questions raised by the reviewer:

a. Are certain event types better suited to the method? -> it is likely that synoptic scale events with minor contributions from mesoscale structures – such as heatwaves for example — are likely more suited for the method because better analogues are likely to be found.

b. Insufficient past data and use of large model ensemble -> we indeed agree with the reviewer and have discussed these points in L347-352

c. Good analogues in later decades -> if such a thing appears then it would point towards an increasing probability of the synoptic pattern of the event. This should not impact the conditional probability statement but would qualitatively suggest an increasing unconditional probability of the event. However, validating such a condition quantitatively is likely difficult (as mentioned in L57-58) and subject to the possibility that the increasing recurrence of this synoptic pattern only arises as a result of natural variability.

Minor comments

1. The 'past' starts in 1950, restricted by the ERA5 dataset. Do you think this matters? Do you think results would differ significantly if you were able to go right back to 1850?

It would definitely be more interesting to start the analysis at an earlier date, simply because it would imply having more data points and therefore improve the quality of the statistical analysis. However, as can be seen in Figure 3abc, the analysis lacks data points for high RMST, i.e. for the present with global warming. It is therefore probably more this period that would benefit from more sampling. Finally, going back to 1850 would allow us to sample more the natural variability of the climate system but, given that we already use 70 years of data, we do not expect to have large gains from going this far in the past. As a consequence, we do not expect that going much farther in the past would significantly change the results obtained.

2. Line 65 – dates formatting, could remove the 'of''s throughout (e.g. 4th July, not 4th of July)

This has been corrected.

3. Line 98 – why do you take Z500 rather than SLP (as used in Faranda et al 2024)?

The choice of using Z500 in this study is justified by previous studies (e.g. Faranda et al. 2020 or Jezequel et al. 2018). We also note that Faranda et al. 2024 use Z500 and SLP alternatively. Using Z500 implies conditioning on the mid-tropospheric circulation while using SLP implies conditioning on the surface circulation. In general, one or the other choice can be made depending on expert knowledge of what is the most relevant for the extreme event studied (or even both choices can be examined to compare the results). Here we chose not to explore the sensitivity of our results to this choice.

4. Line 112 - Some other analogue studies use spatial correlation to identify analogues, why did you choose to use Euclidean distance?

In general, our experience with the use of analogues is that the distance used to find analogues is not a major determinant of the analogues found (Yiou, 2014). As a consequence, we could indeed have used spatial correlation (note that maximizing spatial correlation is the same as minimizing Euclidean distance if the data is spatially normalized). We do not expect to have major differences in the results obtained if we had used spatial correlation to find the analogues (we tested the method using the L1 norm instead of the L2 norm and found no major differences).

5. Line 200 – Assessing linear trend per decade – there aren't many data points, did you test sensitivity shifting the decades (i.e. 1955-1964, 1965-1974 or other starting years)?

We agree with the reviewer that there are not many data points and therefore that the assessment of the trend should be taken with some caution. To test its sensitivity we employed a bootstrap procedure on the analogues rather than shifting the decade as suggested by the reviewer (L209-212). We thank the reviewer for this alternative method but as it is not the core of the results obtained in the paper we prefer to keep the method we proposed.

6. Fig.2 caption – the final sentence is a bit misleading as no trends are shown. I think this should be removed, and just referred to in the results (I spent a while trying to spot the trend!)

This has been removed as suggested by the reviewer.

7. Fig 2 / 3 / 6 – it would be great to title the columns as you do in Fig4

This has been modified as suggested by the reviewer.

8. Fig3/6 could you align the zeros?

We are not sure to understand which zeros the reviewer is referring to. In both panels abc and panels ghi, the range of values and units used are different for each plot, therefore, for ease of visualization we prefer to keep the figure as it is (note that in panels ghi the zeros are indicated by an horizontal line).

9. Lin 255 - 'northern France and 'southern England' rather than north of etc.

This has been corrected.

10. Line 271 – You chose to use the same number of analogues for all events, but some events may have more good analogues (i.e. be less dynamically extreme). Would there be a way to incorporate this into the method, so use a different number of analogues depending how good the analogues are? Maybe by using the information in Fig2a,b,c?

The reviewer raises an interesting point. The choice of the number of analogues we made is indeed somewhat arbitrary (although we tested the sensitivity of our results to this choice). It would be very relevant to have an absolute criteria to define what is a good or bad analogue. Unfortunately we are not aware of a method to do such a procedure systematically. As a consequence, the best we can recommend is probably to proceed as we did by systematically evaluating the sensitivity of the results to the number of analogues chosen. We prefer also to Platzer et al. (2021) for recommendations on the number of analogues for simple models.

11. Fig.4 - z500 contours are not very clear to me

This is indeed the case because for most of them the difference is not significant. Making them more clear to read would imply modifying the results displayed, that is why we had to display them as they are. Although we agree that this is an issue for visualization we prefer to keep them as such with the associated commentary in the text.

12. Line 279 – precipitations (no need for 's')

This has been corrected.

13. Line 345 – stray '?' in references

This has been corrected.

14. Line 392 (final sentence) This doesn't quite read right to me, consider rewording.

We have modified the sentence as follows: L412 " All these attribution statements are conditional to the synoptic scale pattern observed during the events."

References

Jézéquel, A., Yiou, P., & Radanovics, S. (2018). Role of circulation in European heatwaves using flow analogues. *Climate dynamics*, *50*(3), 1145-1159.

Yiou, P. (2014). AnaWEGE: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, *7*(2), 531-543.

Faranda, D., Vrac, M., Yiou, P., Jézéquel, A., & Thao, S. (2020). Changes in future synoptic circulation patterns: consequences for extreme event attribution. *Geophysical Research Letters*, *47*(15), e2020GL088002.

Platzer, P., Yiou, P., Naveau, P., Filipot, J. F., Thiébaut, M., & Tandeo, P. (2021). Probability distributions for analog-to-target distances. *Journal of the Atmospheric Sciences*, *78*(10), 3317-3335.