We thank the editor and the reviewers for considering our work and for their comments. In the following we address their final comments.

Editor's comment

I have received the reviewers' assessment of the revised manuscript and they are overall satisfied with the nice work the authors have done to address their concerns. One reviewer has some rather minor but, in my view, important requests that should not take much time. These final revisions will help make the method/results more interpretable and useful to the community, and ensure that the manuscript aligns well with WCD's aim to advance dynamical understanding of weather and climate.

In addition, I would ask the authors to revisit the description of the quality metric around L125-L128. I believe the confusion stems from some wording problems. I had to go back to Faranda et al. 2022 to understand how Qk is calculated.

Do the authors mean: For each analogue k of an event, the analogue quality Qk is computed as the average Euclidean distance from its 72 closest analogues, which need not a priori be the same as the 72 analogues of the event itself. In other words, the fact that pattern A is among the n = 72 best analogues of pattern B does not necessarily mean that pattern B is among the n =72 best analogues of pattern A.

We have rephrased the whole paragraph as suggested and hope this version is more clear:
L121 "To check the quality of the analogues found with our procedure, as in Faranda et al. (2022), we compute the analogues quality metric Q. For each event, its quality Q is computed as the average Euclidean distance of its n=72 analogues. For each analogue k of each event, its own analogues quality metric Q^k is then computed as the average Euclidean distance from its n=72 closest analogues computed over the same domain and time period as for the event. These need not a priori be the same as the n=72 analogues of the event itself. In other words, the fact that a pattern A is among the n=72 closest analogues of a pattern B does not necessarily mean that pattern B is among the n=72 closest analogues of pattern A. We then compare the value of the analogues quality Q^{\text{event}} for the event with the distribution of the (Q^k)_{1\leq k \leq n} of its analogues. If Q^{\text{event}} is a clear outlier of the distribution of (Q^k)_{1\leq k \leq n} --- for example if it is higher than their maximum --- this means that the synoptic pattern of the event is unique and therefore that it has bad analogues. In this case, a conditional (and probably also unconditional) attribution statement is likely impossible based on past data only."

Also, I find phrases like "All events are in the upper tail of the distribution of analogues quality of their analogues" confusing, particularly "analogues quality of analogues" (if you had to write this, it would be "analogue quality of analogues", but probably still better to avoid it). I feel it would be clearer to talk about the Q for each analogue of an event, the Q distribution for all analogues of an event, and the Q for the peak day of the event, but I leave it to the authors to find a solution they feel works best.

We have rephrased those sentences using the mathematical notation for more clarity:

L207: "We show in Figure 2abc the distributions of Q^k for the analogues found (boxplots) and the analogues quality Q for the event itself (red dot). The analogues quality metrics Q of all events are in the upper tail of the distribution of Q^k --- which may be expected in-so-far as they are all rare events --- but are not outliers of the distributions."

Reviewer #1:

The authors have carefully considered the points I raised in my previous review, and revised their results in a way which strengthens them substantially. The method looks to provide many advantages over the EVT approach. I think the manuscript is now ready for publication subject to one relatively minor revision.

We thank the reviewer for their helpful comments during the review process and answer here their final comments.

Given the likelihood of this technique being used for real societal applications I think it's really important to explicitly show the analogues so meteorologists can understand the strengths and limitations of the attribution in a concrete way. So I'm pushing back slightly on the comment that seeing some analogues would not be very informative. For that reason I would like to see some supplementary 6x12 stamp plots showing the 72 analogues for the temperature and precipitation cases in the style of fig 1 d,f.

We have now added the 72 analogues as supplementary figures for the three events (Figures B1, B2, B3).

Connected to this, a comment on the discussion of figure 1, that might be worth a couple of lines of discussion in section 3.1. You've mentioned that the localised low pressure system is not well captured by the mean analogue pattern, but I think the downstream ridge is also important here, which drives convergence over the precip region. Also the temperature analogue is quite different from the event: just compare the geostrophic flow along the 5850 contour in 1a, and the 5800 contour in 1d. For the event there was strong southerly flow from Spain and perhaps even North Africa. For the analogues the flow seemingly comes quite zonally from the Atlantic. Not surprising from that perspective that thse might be cooler events.

We agree with the reviewer and have modified the text to mention these elements: L200 "For the temperature and wind events the synoptic situation is qualitatively similar to the event itself, and so are the observable fields, although with an intensity that is less than the events as can be expected because we are investigating extreme situations. For the temperature event the mid-tropospheric flow is nonetheless less meridional than for the event. The synoptic situation for the precipitation events is not as satisfying insofar as the composite does not show the secondary trough that can be seen over the eastern Pyrenees in Figure \ref{fig:events}c and the downstream ridge is less intense than for the event. Despite this discrepancy, the structure of the precipitation field is qualitatively similar to the one of the event itself, with precipitations in south-east France and northern Italy."

Reviewer #2:

The manuscript has been updated, based on mine and another reviewers comments - and I approve of the changes. The new manuscript has taken most of our suggestions and comments into consideration. I still feel a few changes are needed to make the paper suitable for publication.

We thank the reviewer for their insightful comments and answer them below.

1. I am still not fully on-board with using the method for precipitation events (or at least the event illustrated) as it seems to give confusing results. For heatwaves there is a much more direct relationship between dynamics and the (temperature) impacts. For rainfall as well as the necessary dynamics a moisture source is also required – and this likely explains why the analogues are not all similar to the event (in terms of the impact – rainfall). It is interesting to show the difference in method performance for different variables, but I still think further discussion and acceptance/explanation of the limitations is required. A start has been made on this around line 395 - but multiple suggestions are made. It would be useful to determine which (e.g. checking how good analogues are, or comparing to a different rainfall extreme in a similar region).

We agree with the reviewer that some care should be taken for precipitation events. We nevertheless prefer to mention both potential explanations: a real dynamical/thermodynamical evolution for those analogues between past and present and simply an inability to sample good analogues. We have modified the text accordingly: L381 "we tend to see no change or a small decrease in the intensity of precipitation conditional on the analogues. This result is surprising and goes against the Clausius-Clapeyron scaling for the strongest precipitation events. However, when looking at the trend in the 2m air temperature for the analogues of this event, we found no significant trend in the region investigated (not shown), which does not come from a change in the seasonality of the analogues found over 1950-2021. This may explain why we observe no significant change and may arise from the fact that analogues of this particular synoptic pattern do not show a significant warming response. This however may also be due to the fact that no or few good analogues of this event exist in reanalysis. The intensity of precipitation events is forced both by synoptic-scale and mesoscale structures: while there may be good analogues of the latter, the former are likely not well sampled in the ERA5 reanalysis. This advocates for a cautious use of the flow analogues attribution method for intense precipitation events, especially when most precipitation are convective."

2. There is a contradiction between line 404 and 416 – in claiming the method is good due to the lack of use of modelling data, and that a useful extension would be to apply the method to model data.

We do not think that this is a contradiction: the advantage of the method is to rely on reanalysis data only but it could be applied also to model data to strengthen the results and/or compare discrepancies. We have modified the text to clarify this point:

L407 “especially to check the consistency of their thermodynamical evolution for observed weather patterns leading to extremes”

3. Almost a continuation of point 1 – but I think further discussion of the cons of the method is needed. The proposed methodology is a great step forward from that currently used by ClimaMeter, but still cannot be applied to all events. Particularly a comment on how the most unusual events may have no analogues would be useful.

We agree with the reviewer and have emphasized this point more strongly:

- **Discussion: L350 “and therefore the strongest, unprecedented extreme events may not be attributable with this method if they have no past analogues.”**
- **Conclusion: L410 “The method nevertheless requires the existence of past good analogues which may not exist in reanalysis data sets for the most intense and unprecedented extremes.”**