

Manuscript egusphere-2025-4976 “A combined storyline-statistical approach for conditional extreme event attribution”

Response to Editor

I have read the revised manuscript and the authors' responses to the reviewers. The authors have made substantial revisions and the reviewers are for the most part satisfied. There remain two main issues (reviewer 1) which are closely related, and have to do with the interpretation of the results.

There is no doubt that reviewer 1 sees the nudged storyline approach as a valuable one for studying extreme event attribution. The reviewer is, however, highlighting that different approaches giving slightly different messages could be meaningful, and that this is in fact a strength of combining approaches. The definition of a "class" of event could thus be seen to depend on how strongly an approach conditions on dynamics (Thompson et al. doi: 10.1002/wea.7779). In addition, for the nudged storyline approach, uncaptured dynamics may still be important (e.g., van Garderen et al., 2020).

A suggestion is that the authors keep the framing of their study as is (e.g., the question posed in L159), as it aligns with published studies such as Feser & Shepherd, but that they remain open to the points raised by reviewer 1 in their presentation of the results. The Thompson et al. report, which is very new, could be a useful additional reference. I don't feel that this will require lengthy text additions, but rather just adjustments to the current text, particularly in the introduction and conclusions/discussion sections, as well in certain statements in the results sections.

Finally, given that the study should appeal to a broad audience, the manuscript could benefit from a round of copyediting and proofreading to make it as clear and readable as possible. English language editing is no longer provided by EGU journals. Some examples for illustration:

- L28: "In contrast, non-heatwave events shaped by complex dynamical processes - in particular when associated with modification of the large-scale atmospheric circulation like windstorms..." - "like windstorms" relates to "non-heatwave events"? sentence needs to be restructured if so.
- L32-33: extra article "either a counterfactual or future warmer worlds"
- L33-35: "counterfactual" should be explained the first time the term is introduced (L33), not the second time (L35)
- Could section 4.3.1 be its own subsection (4.4)? I realize it pertains directly to 4.3, but it is also a wrap-up and contains its own figure, so I think it would be appropriate.

I would like to emphasize that all of us (both reviewers and myself) are very positive about this study, and feel it represents a valuable contribution to the literature. What I

believe reviewer 1 is asking for is a more open and considered interpretation of the results that balances various viewpoints. I'd be happy to help - perhaps in consultation with reviewer 1 - if the authors would like clarifications on any of the points. I would offer reviewer 1 the chance to see the revised manuscript if they are able to provide feedback in a timely manner.

Authors' response:

Dear Editor,

We appreciate your positive feedback and recognition of our contribution to the literature. We note a conceptual disagreement with Reviewer 1 regarding the levels of conditioning in attribution approaches, which we aim to clarify in our revisions. We have revised the manuscript accordingly, with particular attention to refining the research focus and clarifying the main message. We also addressed Reviewer 2's minor comment about adding soil moisture depth in Fig. 8.

In the revised version, we emphasize that we consider both approaches to be equally valuable and do not intend to prioritize one over the other. Our objective is to highlight the advantages of combining them to provide a physical-statistical conditioned framework for extreme event attribution, not to compare them. We have also carefully proofread the manuscript to improve English language clarity and readability. Following your advice, we have renumbered Section 4.3.1 as Section 4.4. Finally, we have reformulated the conclusions to present them in a more balanced and open manner.

In addition, we have incorporated recent studies supporting the use of conditional attribution methodologies and multi-method assessments. Specifically:

Line 30: ...where conditional attribution methodologies play an important role in understanding the event under similar atmospheric dynamics (van Garderen and León-FonFay, 2026)

Lines 38-40: Because each approach captures different aspects of human influence, multi-method attribution is necessary for a comprehensive assessment (Thompson et al., 2026; Barriopedro et al., 2025).

Please find attached our point-to-point responses to reviewer 1.

Kind regards,

Dalena,

on behalf of all co-authors.

References

van Garderen, L. and León-FonFay, D.: The essential role of conditional attribution in understanding complex extreme weather, *Nature Communications*, 17, 1539, <https://doi.org/10.1038/s41467-026-69056-1>, 2026.

Thompson, V., Ermis, S., and Athanase, M.: The need for multi-method extreme event attribution, *Weather*, 81, 40–45, <https://doi.org/10.1002/wea.7779>, 2026.

Barriopedro, D., Jiménez-Esteve, B., Collazo, S., Garrido-Perez, J. M., Johnson, J. E., and García-Herrera, R.: A Multimethod Attribution Analysis of Spain's 2024 Extreme Precipitation Event, *Bulletin of the American Meteorological Society*, 106, E2440 – E2460, <https://doi.org/10.1175/BAMS-D-25-0049.1>, 2025.

Manuscript egosphere-2025-4976 “A combined storyline-statistical approach for conditional extreme event attribution”

Response to Reviewer 1

I thank the authors for their extensive response to my and the other reviewer's comments. In my opinion the manuscript has improved substantially, especially when it comes to addressing our remarks on the interaction with soil moisture when comparing the two approaches. The new analyses and figures provided give relevant information for the readers. That being said, it still seems to me that some of the key claims made by the authors are not supported by their analysis, rather that their new figures support my earlier points about the change in the probability for the heatwaves. The attribution literature needs more comparison studies between different methods, and I think the results presented here are a very positive development in this direction, therefore I still support their publication in WCD. However, the main claim of the authors that a similar atmospheric pattern leads to much increased probabilities in a warmer climate and the associated implicit assumption that the nudging approach gives the “true” intensity change is not, in my opinion, supported by the results presented and therefore I still require major revisions. I would not oppose the publication of the results, presented as a discrepancy between the two methods, but I think the authors claim something much stronger than that. Please find below my main points.

Major comments

1. Attribution of a single event: This point is a little bit epistemological but I wanted to start with it because I think it has consequences for the whole study. The authors claim in their response to my previous comments and in the manuscript that the nudging method do not create a new class of events but rather addresses the peculiarities of a particular event. As I argued previously, I strongly disagree with this statement: we cannot make any attribution statement about a single point in the phase space, what we compare is always, implicitly or explicitly, a group of events in different periods which bear similarities with the one observed. Thus, the question posed in L159 (“What is the influence of anthropogenic global warming on the extreme event of study?”) is either too general (this is what every attribution method seeks to achieve) or meaningless (in the sense that we cannot give a precise mathematical meaning to what is the quantity measured). The nudging method strongly constrains the dynamics, but there are still degrees of freedom: this is the variability seen in Figure 2 or Figure 8a between the different nudged members. Therefore, the question is always which group of events you compare to and whether that is relevant or not for the question asked.

Authors' response:

We understand your point. From a strict mathematical perspective, this is correct. We are aware that even storylines could be a ‘class of events’ but that would be a very restricted one. This is just a common terminology in attribution. Depending on the level

of conditioning, highly conditioned approaches like Spectrally Nudged Storylines or Pseudo Global Warming are said to study specific events based on prescribed dynamics, conditioned approaches like flow-analogues study events under a class of similar atmospheric dynamics, and unconditional approaches like WWA attribute a class of events coinciding with the same magnitude as the event of interest but grouping events deriving from different dynamical situations (Thompson et al., 2026).

2. Changing probabilities: I am not convinced by the response of the authors when it comes to why events with similar dynamics in the future become much more likely. I argued in my previous comment that the nudging members have a different soil moisture dynamics than the analogues because of the way the nudged simulations are made. Thanks to the precisions given by the authors and their new Figure 8, I am now much more certain of my interpretation:

- the authors say L123: "Each storyline represents a physically consistent, plausible scenario in which our observed climate could have developed under an alternative thermodynamic background" and I think this statement is not true. The nudged storylines represent 10-year long simulations done by imposing GHGs and SSTs to an atmospheric model that is additionally nudged above 750hPa: imposing both the thermal gradients and the circulation breaks the coupling between the dynamics and the thermodynamics, as expressed for example by the thermal wind relationship. Thus, at least according to the physical laws we know, the atmospheric circulation in a warmer/colder world could not have been the one observed during these 10 years. - This implies in particular that the preceding soil moisture distribution before the event is not the climatological one: it incorporates only the forced thermodynamical trend, not the dynamical one. On the contrary, the analogues do have both thermodynamical and dynamical trends. This explains why in your new Figure 8b the number of analogues with a soil moisture anomaly greater than the one of your nudged events is increasing with increasing global warming. This is evidence that you compare events which are drawn from different distributions, which implies that you cannot project the probability as proposed because these events are fundamentally not occurring in the same climate (in addition to the fact that the model setup in the two cases is different). In my opinion this is also evidence that the nudging approach underestimates the change in the heat extremes because of a lack of correct projection in the soil moisture statistics. It seems totally okay to me that the two approaches disagree, especially because the authors have a physical explanation for why this is the case. I agree that the authors did their best to compare the methods and this is not what I am criticizing. As soon as one acknowledges the differences, advantages and disadvantages of the two methods, two choices are possible: either comparing and trying to explain the differences, or saying that one is better (in one sense or another) to the other. The second choice needs better arguments than the first one, and it seems to me that the authors took the second option without sufficient evidence for it.

Authors' response:

We thank the reviewer for the constructive comments. However, we respectfully disagree with the interpretation that the two approaches sample different climates. The core of the disagreement appears to derive from a misunderstanding of the fundamentals of conditioning in attribution methods, which was previously clarified. Storylines are by definition conditional experiments. They impose a prescribed dynamical state while modifying thermodynamic boundary conditions. They isolate the thermodynamic response along a single dynamical pathway under alternative warming levels. In the revised manuscript we've explicitly stated that they are hypothetical and should not be regarded as future projections. In contrast, the flow-analogue approach samples multiple dynamical realizations at the same warming levels. The difference lies in the degree of dynamical conditioning, not in the underlying climate state. The storyline constrains dynamics strongly and explores thermodynamic amplification of a single event, while the analogue framework captures a broader range of antecedent and interacting processes (soil moisture availability is only an example). However, we agree that in the flow-analogues distribution, decreased soil moisture could potentially induce an increase in temperature variance leading to more extreme heatwaves.

Lines 59-64: While the method successfully evaluates changes in the characteristics of individual events...**it may not account for events that remain dynamically consistent with a warmer climate, in which the background atmospheric circulation may shift (Thompson et al., 2026). This raises a fundamental question within this combined framework: How likely is this weather situation, and how plausible are the projected storylines under warmer climate conditions?**

Lines 459-466: Given that in the storyline approach we deliberately focus on a single dynamical pathway, even if it wouldn't happen in the same way in alternative climates, we obtain the thermodynamical response on a single event. With the flow-analogue method, we identify that even more extreme heatwaves could emerge under similar atmospheric circulation, but different preceding and prevailing conditions to the 2018 heatwave in a warmer climate. This result must be interpreted considering the distinct purposes of the two approaches. Because storylines are hypothetical experiments, they cannot be regarded as projections or claimed to happen exactly as projected in the future. However, flow-analogues provide the statistical and dynamical context to address the plausibility of similar large-scale circulation and events emerging with comparable intensity under future climate states.

We also acknowledge that unintended priority was given to the storyline approach initially. We have revised the text to emphasize that the two methodologies are complementary rather than competitive. Storylines quantify thermodynamic intensification along a prescribed pathway, whereas flow analogues provide the statistical and dynamical context to assess how frequently similar circulation

patterns and comparable or greater intensity emerge under warming. In the revised manuscript, we have edited these statements in a more balanced way. We also reformulated the main conclusion from “storyline-projected intensities become more frequent” to “heatwaves occurring under similar atmospheric conditions, and exceeding the storyline-projected intensities, may become more frequent and extreme...” This modification is subtle but conceptually important. It clarifies that we do not claim an increasing recurrence of specific storyline events (or claim that they will happen exactly in the future). Instead, we refer to the increasing occurrence of heatwaves developing under comparable large-scale circulation patterns and reaching equal or greater intensities under warming.

Lines 5-6: “Despite no detected changes in terms of atmospheric blocking by the flow-analogue approach, **the combined framework further indicates that heatwaves under similar atmospheric conditions, and exceeding the storyline-projected intensities, may become more frequent and extreme...**”

Lines 73-76: **In our case, we use it to identify dynamical changes by quantifying shifts in the frequency of good analogues with global warming, and to assess the probability of comparable events.** While this method is less suited to attributing the precise influence of anthropogenic warming on a single event, it provides robust estimates of the likelihood of similar events occurring under varying climate conditions, important for risk assessments (Trenberth et al., 2015).

Lines 102-104: In this way, this framework aims to complement the individual strengths of each method by jointly addressing thermodynamic event intensification and circulation-dependent probability changes with global warming.

Lines 406: Heatwaves occurring under analogous atmospheric circulation, and with equal or greater intensity than those projected by future storylines, increase in frequency with global warming.

Line 467: Together, these two perspectives provide complementary insights...

Line 471: The flow-analogue method assesses the likelihood of similar atmospheric dynamics, and the heatwaves derived from them, in present and alternative climates.

Lines 482: Importantly, this conclusion is event-specific. For other cases, future dynamics could diverge sufficiently so that warmer versions of a historical event may become physically implausible. Using both approaches in combination allows us to identify when this is the case.

Minor comments

1. L45: “The traditional risk-based approach, like the World Weather Attribution method (Philip et al., 2020), would not align with our purpose of focusing on a specific event due to its unconditioned nature” → I am not sure I understand this statement: every extreme event attribution method seeks to address the attributability of a specific event, this is what every study of the WWA does. Same point for the sentence L75.

Authors' response:

We've removed this statement to avoid misunderstandings.

2. L82 “Conversely, the flow-analogue method, based on large ensembles, offers probabilistic estimates of risk while lacking the event-specific physical detail of the storyline approach” → I am not sure I understand this statement either: the point of conditioning on the analogues is to condition on the dynamics, thus on similar physical mechanisms.

Authors' response:

The flow analogues are only conditioned on the dynamics; they have the same circulation pattern but differ in other preceding and prevailing physical mechanisms. Storylines are highly conditioned chronologically on the same dynamical pathway as the observed event, so that is event-specific.

3. L147: “Due to the high-frequency temporal output (3 to 6 hourly) availability for its 50 realizations, the model provides a sample size large enough to work with flow-analogues and extreme events” → It seems to me the improved sampling comes from the 50 members rather than the high-frequency output, because you also impose that analogues must not be too close.

Authors' response:

We agree, thank you. The sentence has been reformulated to reflect that the ensemble size is the main factor enabling robust analogue sampling, even though a high temporal (at least a daily) resolution is still important to be able to accurately simulate the daily analogue sequences.

Line 147: Due to the availability of 50 ensemble members at a sub-daily resolution...

4. Paragraph L149: just as a precision, are members treated independently or do you do that on the forced component estimated via the ensemble mean?

Authors' response:

Members are treated independently for precision.

5. I thank the authors for their detailed explanation of how the analogues are found but I think this should be incorporated in the manuscript. For example in L166: how are analogues detrended and detrended with respect to what ?

Authors' response:

Line 164: Analogous circulation patterns are identified at different global warming levels from the MPI-ESM-LR Grand Ensemble under SSP5-8.5. For each member and warming level, GPH anomalies are computed relative to the 1985–2014 mean climatology, linearly detrended within the warming level period, and restricted to summer months (June–August).

6. L217: I still do not see why a GEV distribution would be suited for the conditional distribution of temperature. A GEV distribution is a very peculiar distribution, which is bounded upwards or downwards depending on the value of the shape parameter: this value determines what is happening far in the tail and is the key parameter to estimate for extrapolating (the location and scale parameter only center and scale the distribution). Thus we need good evidence for why to use this distribution, which is not provided here. The authors argue in their response that they use a GEV to “account for uncertainty ranges using bootstrap-resampling and to extrapolate magnitudes and return periods that go beyond the available ones in our dataset.”, but this is exactly the point: how do you know that the GEV can extrapolate beyond the available data?

Authors' response:

Thanks for your comment. We noticed we made a mistake in our previous response, since we are not actually extrapolating but only interpolating to obtain fitted values within a range of uncertainty. In Figure 7, we plotted the flow analogue data before fitting to show that GEV represents very well the dataset. Here we can also see that the fitting doesn't go beyond the analogues' available data.

7. L455: “Yiou et al. (2020), who identified no dynamical trends for good analogues of the 2018 heatwave” → please note that this is a model result that, given recent works on discrepancies in the dynamical response of models, we should be careful with

Authors' response:

Thank you, we have pointed this out.

“... of the 2018 heatwave in **CMIP5 RCP4.5 and RCP8.5 simulations.**”

8. L483-485: I have difficulties understanding the specificity of this sentence: every attribution method is trying to answer this question.

Authors' response:

In a certain way it is, but this is again related to the different purposes of each attribution methodology. We have removed this statement.

9. L507: “Even within such recurrent circulation patterns, the 2018 heatwave in a factual world was still very unlikely.” → maybe you should precise that this statement refers to the combination of the temperature anomaly and the atmospheric pattern, not only the temperature anomaly

Authors’ response:

Thanks, we have clarified this point.

Line 475: Even within the recurrence of a blocking over Scandinavia, a heatwave as intense as the 2018 heatwave in a factual world was still very unlikely in present times.

- 10.L512: “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” → as explained above, I am skeptical of this conclusion.

Authors’ response:

We have rewritten this statement.

Lines 482: Importantly, this conclusion is event-specific. For other cases, future dynamics could diverge sufficiently that warmer versions of a historical event may become physically implausible. Using both approaches in combination allows us to identify when this is the case.

References

van Garderen, L. and León-FonFay, D.: The essential role of conditional attribution in understanding complex extreme weather, *Nature Communications*, 17, 1539, <https://doi.org/10.1038/s41467-026-69056-1>, 2026.

Thompson, V., Ermis, S., and Athanase, M.: The need for multi-method extreme event attribution, *Weather*, 81, 40–45, <https://doi.org/10.1002/wea.7779>, 2026.

Barriopedro, D., Jiménez-Esteve, B., Collazo, S., Garrido-Perez, J. M., Johnson, J. E., and García-Herrera, R.: A Multimethod Attribution Analysis of Spain’s 2024 Extreme Precipitation Event, *Bulletin of the American Meteorological Society*, 106, E2440 – E2460, <https://doi.org/10.1175/BAMS-D-25-0049.1>, 2025.

Trenberth, K. E., Fasullo, J. T., and Shepherd, T. G.: Attribution of climate extreme events, *Nature climate change*, 5, 725–730, <https://doi.org/10.1038/nclimate2657>, 2015.