Review of "A combined storyline-statistical approach for conditional extreme event attribution"

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 constrain 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.

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.

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.

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.

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.

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 ?

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 ?

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?

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

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

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

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.