

## **Review of “A combined storyline-statistical approach for conditional extreme event attribution”**

The work of León-Fonfay et al. proposes a framework to combine two conditional attribution methods of extreme events, one based on nudged simulations and the other one on flow analogues. While the first one gives thermodynamical changes for a dynamics following very closely the actual event, the other one follows more loosely the event but is able to assess conditional probabilities. The authors propose to combine the two to recover probabilities associated with the nudged approach. The method is illustrated by applying it to the 2018 Central European heatwave.

The paper is well-written and proposes an interesting method for improving results on an important topic of attribution. I especially appreciated the effort to find analogues in a large ensemble, which clearly enhances the quality of the results with this method. It is also a very worthy point to compare and combine attribution methods. It is therefore, in my opinion, suited for publication in WCD. I found the results obtained by the authors really interesting but also quite surprising and puzzling. I am not sure I agree at this point with their interpretation, especially on the fact that events such as the 2018 heatwave become much more likely in their own climate, including after conditioning on the same atmospheric dynamics. It seems to me that it may as well illustrate issues of the nudging method. I think this point needs more investigation to give an idea why the same dynamics, while being correctly recovered by the analogues, can give rise to much more intense events in the future. I therefore recommend major revisions. Please find below the details of my comments.

### **Major comments**

1. L125: The storylines simulation are nudged using NCEP reanalysis but you use ERA5 reanalysis to find the analogues. This seems like a potential issue to me. I do not expect the two reanalyses to have major differences but there could still be some. Especially the discrepancies in the temporal evolution in Fig 2c between ERA5 and the factual storyline do not seem so small. Please discuss that thoroughly or change the analogues analysis towards using NCEP data for the event.
2. Sec 3.2: I think the way analogs are found needs some precision because this may be important for interpreting your results correctly.
  - a. How are the GPH anomalies detrended precisely ? Please note that using grid point anomalies and detrending by grid points break the horizontal gradients and therefore the closeness between the winds of the event and the analogues (because of the geostrophic balance). In particular the model and ERA5 may not have the same climatology and therefore comparing anomalies may not compare patterns that are actually physically similar. Similarly, the climatology likely changes between the different periods in the model (both in mean and variance). Finally, the climatology of ERA5 is problematic to compute: how did you estimate the forced component in GPH ?
  - b. Am I right if I say that you compute the euclidean distance over 5-days or are you averaging first over these 5 days ? The first option seems problematic to me because it vastly increases the dimension of the space in which you are

looking for nearest neighbors and therefore risks finding worse analogs (in other words, it is way more difficult to find events which have the same dynamics over 5 days than over 1 day).

- c. I find the region to look for the analogues slightly strange: you are interested in temperature in Germany, therefore the anticyclone (which is local) over Germany should be your target field. I am not really convinced about using a box so extended to the West. In particular, it is possible that in your analogs you have events with a low pressure in the south of Iceland but few high pressure over Germany. That being said I am not sure it would be a major change to the results because you already have a large part of the domain included.
  - d. I am not sure I understand your two steps procedure to find the analogs: why first selecting the  $N$  closest events if after that you impose again a hard threshold on the spatial correlation? Why not directly using the spatial correlation? By the way I think using a hard threshold of spatial correlation requires some thinking: the correlation for each day is scaled by the spatial variance of that day, which would be different from one pattern to another. This implies that events with a similar absolute distance may have different spatial correlation. This has important implications when estimating  $P(D)$ . I would advise rather using a similar absolute euclidean distance for each period because this has a straightforward interpretation: this is a ball around your event of interest in the phase space and counting the number of events in the ball gives you an estimate of the probability to fall into this ball, i.e.  $P(D)$ . Estimating the sensitivity of your results to this threshold would also be good given that this is one of your main results.
3. Sec 3.4: I am not sure I see the interest of computing the probability  $P(E,D)$ . I think there are two interesting probabilities here: either  $P(E|D)$  which is the probability to observe  $E$  given the pattern  $D$  — the statement in L192 that “the probability of an event of magnitude  $M$ , under the circulation pattern of interest  $D$ , in a given global warming level (GWL) will be defined as  $P(E,D)$ ” is not true I think —, or  $P(E)$  which is the unconditional probability to observe your event. Because  $P(E) = P(D)P(E|D) + P(\sim D)P(E|\sim D)$  you could compute this absolute probability, but this requires to know  $P(E|\sim D)$  which is a probability that can change with climate change. Thus I am not sure that an absence of change in  $P(D)$  can be so straightforwardly interpreted. Finally, you mention in L195 that you fit a GEV to the analogs conditional distribution and I see no reason to proceed as such: the use of GEV in extreme value theory is justified by the extremal theorems that state that the limit of block maxima distributions should follow a GEV. You are not using block maxima so it is not clear why your conditional distributions should follow a GEV (Figure 6 in particular does not plead in your favor). In a similar context Noyelle et al. (2025) used for example a skew normal distribution. Given the results of your Figure 6 and the large number of analogs you have, why not simply computing the probability empirically by counting the number of analogs above your target?
4. The most surprising result of this paper is that even when compared to their own climate, the probability to observe an event like the 2018 one increases vastly. This result is obtained by comparing (i) the intensity of the 2018 event obtained by the nudged storyline method in different climates and (ii) the probability to observe such an intensity in the new climate conditional on having the same large scale

atmospheric pattern (i.e. using the analogs). From this, Figure 6 clearly shows that this latter probability increases in future climates because of an increase in the variance of the conditional distribution. This increase in the variance is expected and likely comes mainly from the strengthened feedback with soil moisture in a hotter, drier summer climate in Europe (e.g. Fischer et al. 2012), because, as shown by the authors, the intensity of the analogs GPH pattern (and therefore by extension of advective and adiabatic contributions of heatwaves) does not seem to change. On the other hand the change in the mean seems to be mostly linear with GW, which is also reflected in your nudged simulations (Fig. 4b). The authors then interpret this change as a real change in probability for such an event as in 2018 and this is where I am not convinced. I think the way the nudging simulations are designed strongly constrains the soil moisture component and therefore that the method proposed here essentially compares a distribution with fixed soil moisture (nudging) and one with varying soil moisture (analogues), which inflates artificially the probability obtained. To detail: the nudged simulations run for 10 years with the winds observed over the period 2015-2025 in climates with different warming levels. This implies that the soil moisture component is driven by different thermodynamical forcings but similar dynamical ones, which is a strong constraint given that rain and evapotranspiration are strongly dynamically driven variables. As such, the soil moisture distribution in the nudged simulations — and especially at the beginning of the 2018 event — are likely different from the one of the MPI LE, from where the analogues are taken. For a fair comparison, the winds should be nudged for a realistic distribution of soil moisture (i.e. with respect to the soil moisture distribution of the LE) some weeks before the event or the analogues should be taken conditional on a soil moisture distribution similar to the nudged simulation (which would likely decrease the variance in Figure 6 and therefore the change in probability). Here you assume that the change in intensity given by the nudging is correct and infer from that the change in probability from the analogues. I think one could do exactly the inverse: given the probability in the factual period, take similarly unlikely events in other periods and infer from that the change in intensity, which would rather show that the nudging underestimate the change in the intensity of the event. Currently, I am personally more convinced by this interpretation than by yours.

To take a step back and summarize, my main issue with the current reasoning presented is that the authors assume that nudging the winds in a different climate only conditions the dynamical component: this is not true in general because global warming has both dynamical and thermodynamical components that are intrinsically linked. One main example is the thermal wind relationship: by imposing the winds you also impose the thermal gradients, despite the fact that those thermal gradients do change with climate change. In other words, the nudged simulations impose an atmospheric dynamic that is physically not realistic — at least at long time scales.

### **Minor comments**

1. L13: please note that cold waves have decreased worldwide
2. L59-61: it is not clear to me what the numbers are referring to precisely, would it be possible to give more precision of what are the events considered and what is the climatology here ?

3. L73: “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” -> I am not sure this is true: the ‘precise influence of anthropogenic warming on a single event’ is not something that is well defined because a single event is *single*, i.e. it will never reproduce identically in the future, even in a stationary climate. All attribution methods require creating a class of events. This is also true for the nudging method: this is the class of events that have exactly the same upper-level winds in the future as the event observed. What is at stake when comparing attribution methods is precisely to know how the class of events defined gives a relevant answer to the question one wants to answer. In this sentence you seem to assume that you know which method is a priori better, which is really not sure I think.
4. Sec 2.2: It was initially a little bit confusing to me what you call storylines in this section, because you seem to argue that a climatological simulation of a +2K world is a storyline. In the end you use the nudged simulations only for some days around your event, maybe you could then present the nudged simulations in a simpler way. Also, could you precise at which levels the winds were nudged ?
5. L149-155: I think I disagree with the statements in this paragraph. I do not think it is true to say that the nudged storylines answer the question “what is the influence of anthropogenic global warming on the extreme event of study”. I must say I think this question is not well posed and therefore cannot be answered without additional precisions: if you consider the phase space of the climate system and you take a point in this phase space (i.e. your extreme event), it is meaningless to ask how climate change changed this point. An event is an event, it does not mean anything to ask how it changed. You can ask how its probability changed (change in the stationary distribution for this point), how its intensity changed (change in the probability to observe a function of this point) etc but these questions always require creating a *class* of events, i.e. a set in the phase space. The discussion about different attribution methods is precisely how to define these classes so that they answer the question we want to answer. The nudged storylines method is just one way to define this class, and maybe not the best (see my major comment 4).
6. Fig 3 and L216: at these places and several other places you do not define precisely what is the climatological percentile that you compare the storylines to. If it is the “local 95th percentile of the climatological daily maximum temperature” then I think the figure is wrong because the local percentile should be the percentile of the climate of each storyline. I guess you are comparing to the percentile in the factual climate, which is fine, but this needs to be precised so that the reader understands that most of the change comes from an increase in the mean. Same comment for the climatological 95th percentile in L286.
7. Fig 4a: Is this a mix of all storylines ?
8. L314: “In a factual world, the 2018 heatwave conditioned to such an atmospheric circulation pattern had a 1-in-277-years” -> I don’t think this is correct, the conditional probability/return time is the one you give at the next line:  $1/1.6\%=62.5$  years.

## References

Noyelle, R., Faranda, D., Robin, Y., Vrac, M., & Yiou, P. (2025). Attributing the occurrence and intensity of extreme events with the flow analogue method. *Weather and Climate Dynamics*, 6(3), 817-839.

Fischer, E. M., Rajczak, J., & Schär, C. (2012). Changes in European summer temperature variability revisited. *Geophysical Research Letters*, 39(19).