Reviewer 1

In this paper, Raymond et al, 2023 disentangle the relationship between Atmospheric Rivers (ARs) and humid heat occurrence over different regions in the US during the warm season (May to September). This approach considers peaks of wet-bulb temperature and computes the probability of occurring when detecting an AR at grid point and region level. This is also done for two other variables that typically represent ARs: precipitation and Integrated Water Vapour (IVT). Moreover, composites before and after the peak are calculated for key quantities related to humid-heat and ARs, allowing us to infer from the statistical relation which processes are key for these events.

I find this paper very interesting. It was quite easy to understand and enjoyable to read. The proposed method to assess such interaction between humid-heat and ARs focuses on the peak of humid-heat in order to examine the processes that cause humid-heat extremes rather than their maintenance. My main issue was with the "extent of the Data and methods". I was expecting a little bit more of explanation (e.g. selection of humid-heat days methodology, more detailed explanation of Relative Risk metric). There are some aspects that the authors should address before this paper can be published in NHESS. I will list them here, together with some minor/technical corrections.

We would like to thank the Reviewer for their efforts in summarizing our study and recommending improvements to it. We have included more detailed description of the methods, as well as made various other changes. Please find below further details on what we have changed and added (text in red). Quoted reviewer commenters are in *italics*, and our responses are in bold.

L10-11: *Consider rewriting the second sentence of the abstract to improve readability.*

We have trimmed several words to make the sentence simpler:
"Process-linked connections between these two extremes, particularly those which cause them to occur close together in space or time, are of special concern for impacts."

Section 2.2 (L64-69): *What is the domain used for the AR-detection algorithm? The algorithm has geometric criteria, does it have a minimum AR extension threshold? If the domain where the detection algorithm is applied only considers the continental US, this domain can miss a significant amount of ARs due to its geometric criteria. Moreover, it can affect the results, especially in the Western US or areas close to the boundaries. If the AR detection domain used does not take this into account, consider applying the AR detection algorithm in a larger domain where the geometric criteria will not prevent detecting all the ARs. What one could do is to check if the number of detected ARs is similar to the AR Catalogue by Guan [1].*

The AR criteria include an extent of 2000 km in length, and a length/width ratio >2, in addition to several IVT requirements as detailed in Guan & Waliser 2019. Our language was perhaps unclear; we simply look through the Guan-Waliser AR catalogue for AR

presence/absence at each gridcell for each timestep. The catalogue was produced at a global scale, so for example an AR day at a gridcell in northern Washington State can be associated with an AR that is primarily located outside our analysis domain, such as over the ocean or British Columbia. In other words, the US domain that we use does not restrict the ARs included — all portions of ARs that affect US points are included in our analysis.

We have modified the text to make this catalogue usage clearer:
"Using the Guan-Waliser AR catalogue, we subsequently define AR gridcell-days as those for which an AR is present at a gridcell for at least two of that day's four timesteps. The entire AR need not fall within the US domain, as the catalogue is defined globally and we evaluate AR occurrence gridcell-by-gridcell."

*In **Section 2.3** it is explained how peak humid-heat days are selected. This method seems very restrictive, but is justified to explore the processes that lead to cause the humid-heat rather than maintaining it. Despite that, it would be good to know if the results are sensitive to the thresholds used for the "peak" framing or without the "peak" framing at all; consider including the sensitivity tests for this methodology if you have them.*

This is a good idea. The new Figure A12 (see below) presents the same analysis as in Figure 5 but using all humid-heat days (i.e. all days with wet-bulb temperature exceeding the 95th percentile). Comparing the two figures reveals that the peak definition has little impact on the core results or their interpretation. We retain the original Figure 5, however, to ensure that we focus on processes leading to the build-up of humid heat rather than sustaining it.

Also of note, we also have now included a figure (Figure A1) illustrating the "peak days" method, in response to a suggestion from Reviewer 2.

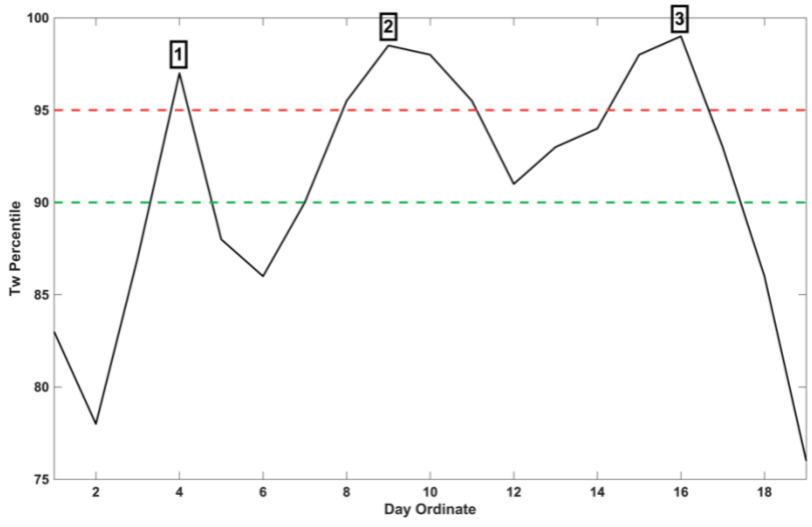
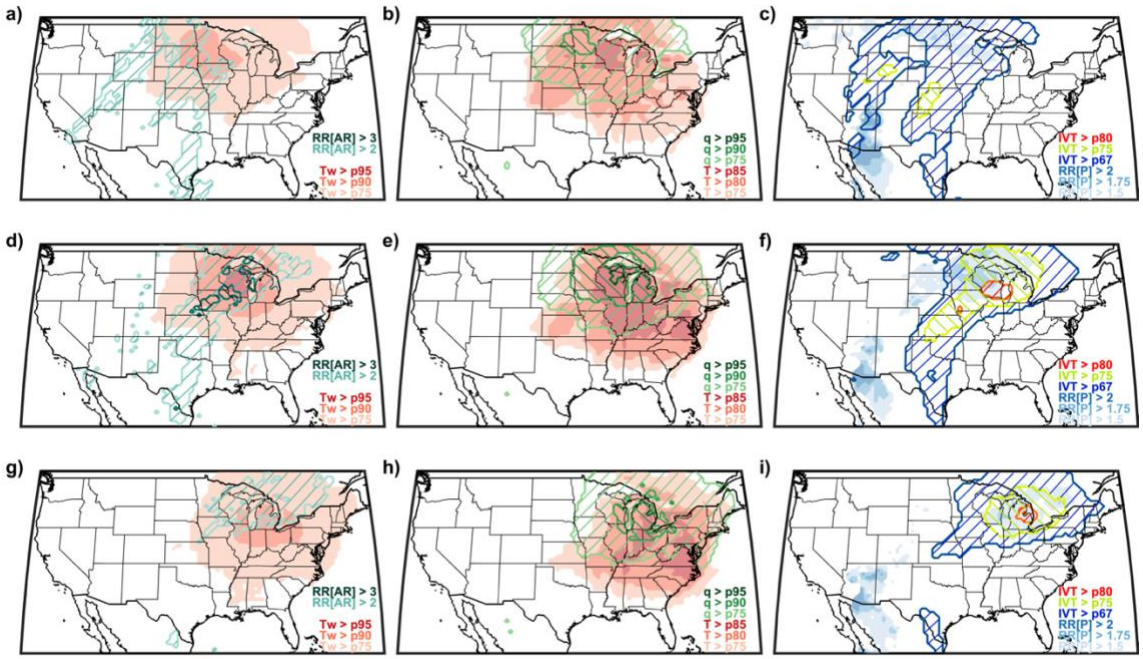


Figure A1: Illustration of the definition of a peak heat-stress day. Days marked 1 and 2 satisfy the requirements of having the highest Tw value within 3 days on either side, as well as Tw having been below the 90th percentile within the preceding 3 days, while day 3 does not. As stated in the text, these requirements can apply to data from an individual gridcell or to a regional (spatial) mean.

Figure A12: As in Figure 5 but for all humid-heat days.



***L94-95**: the Tw percentiles are computed over a 30-day-smoothed climatology. Why is this smoothing necessary to compute the 95th percentile? Then, to define if a day is above the 95th percentile from the smoothed climatology, you use the daily values without smoothing, which percentage of the total days fall above this threshold? Moreover, a 30-day smoothing seems too strong, can you justify why 30 days and not 7 or 15 days, for example?*

Our text was unclear, and actually the smoothing (which we had computed at an earlier stage of the analysis) does not factor into our final percentile calculation. The revised text has been simplified and corrected to read as follows:
"We compute Tw percentiles for each day at each gridcell against the climatology of the surrounding 30 days, then define a 'humid-heat day' as a day with Tw above the 95th percentile."

***L121-122:** What do the colours in brackets mean in the section a) of the caption for Figure 2? I thought the colours represented different US regions, it is confusing, consider removing them if they do not provide any useful information.*

Yes, the colors in the figure represent different regions; the colors in parentheses in the caption erroneously referred to a previous version of the figure, and have been removed.

L122-123: *The caption for Figure 2 section b) is not clear. Seems confusing when compared to what is written in section 2.4 (L132-135). Consider writing here what is shown in the Figure, but also make sure that it is consistent with section 2.4 (L132-135).*

Thank you for noting this. Upon review, we realized that the caption title was also slightly confusing and unrepresentative. The corrected caption reads:
"Figure 2: Relative risk of humid heat by AR intensity and extent
a) For each region, the relative risk of a humid-heat day that has no AR within 1 day and 100 km ("nearby"); with an AR of category 1-3 nearby; and with an AR of category 4-5 nearby. b) Relative risk of humid heat, normalised by regional Z500 anomalies (see Methods), for different AR extents. Note that most regions lack any days with >80% regional AR coverage."

We also revisited the section 2.4 text and revised it as follows:
"As an additional metric for assessing how ARs and humid heat are connected, we compare two sets of days: one comprising all regional-humid-heat days, the other comprising a random selection of non-regional-humid-heat warm-season days with identical regional-mean 500-hPa geopotential height [Z500] anomalies. In other words, normalised by Z500 anomalies, we ask whether days with larger AR extents are more likely to experience humid heat within one day before or after."

L131-132*: Relative Risk metric is introduced and widely used in the paper, consider going more in-depth with the explanation of how it is calculated. And if possible, add a reference as well, this will facilitate interpretation of the results, especially for people not used to it. I would mention that "the particular sets of days" correspond to the peak humid-heat days, as this selection of days is always used in the calculation of relative risk. You could mention here that this is not only done for AR/humid-heat, but also for the precipitation threshold (1 mm) and IVT (Figure 4). Specify, which thresholds are used for these two variables, the precipitation threshold is described in Figure 4 caption, but for IVT has not been explained.*

We agree that the previous text was somewhat vague and have revised and expanded it:
"Relative risk in general refers to the risk of an event of interest in a certain case relative to its risk in a control case; here, it refers to the computed probability of ARs near peak humid-heat days (i.e., of AR/humid-heat interaction) versus the probability which would be expected if ARs and humid heat were randomly distributed relative to one another throughout the warm season."

We have added a sentence mentioning that we do the analogous computation for precipitation and IVT, including the thresholds used:
"We analogously compute relative risk for precipitation/humid heat and IVT/humid heat, using the thresholds of 1 mm/day for precipitation and the local 75th percentile for IVT."

This IVT threshold is now mentioned in the Figure 4 caption as well.

L149/L223: *Relative likelihood is relative risk? Consider using one terminology or introduce this term in the methods section 2.4 (L131).*

We have now standardised the wording by changing all instances of 'relative likelihood' to 'relative risk' and of 'likelihood' to 'probability', as we agree that using multiple and related terms for these concepts is unnecessarily confusing.

L149: *Why the AR relative likelihood is within 2 days? In methods, Section 2.4 (L129) is stated to be within 1 day.*

Thanks -- this was another outdated caption, and has now been fixed.

L167: *change 500-mb to 500 hPa, as it is in the Methods.*

Corrected.

L172: *The results in the western regions can be sensitive to the AR detecting algorithm as mentioned in my comment for Section 2.2. Please, make sure your results are not limited by this issue.*

We agree that this is an important issue, but due to our usage of the globally defined Guan and Waliser catalogue (as described above), our analysis is not geographically limited. We have double-checked and confirmed that there are no artificial constraints of this sort introduced in our code — for each gridcell, all ARs that affect it are included.

L194: *Which is the threshold of IVT used for the IVT/humid-heat relative risk? Consider writing this in the caption of the figure, but also in the Methods section.*

As stated above, we have added to section 2.4 that we use the 75th percentile of IVT for the relative-risk calculation, and we also state this in the Figure 4 caption.

L200: *What is total IVT? Could you explain how it is computed? I have seen total IVT as the integration of an IVT vertical section across an AR to calculate the total amount of moisture an AR transports. I assume here this is not the case, please explain what it stands for.*

'Total IVT' is a term taken from the Guan-Waliser AR catalogue and refers to the sum of the (vertically integrated) north-south and east-west components. Because this is more of a technical detail, we have dropped the adjective 'total' for clarity, and simply call it 'IVT'.

L207-209: *Here you state: "the decrease of dry-bulb temperature due to the shifting position of the ridge-trough system causes maximum Tw to occur on the first day of the pair", the Tw occur the first day of the pair, because the methodology on selecting peak Tw days forces to be this*

way. You could say that the day after the Tw peak (or the second day of the pair) the dry- bulb temperature decreases due to the shifting position of the ridge-trough, but the specific humidity (q) remains as high as the first day of the pair (or Tw peak day). I think you cannot imply causality in this case.

Indeed, our phrasing did not properly account for the complete methodology, and especially the fact that Tw peaks on the central day of Figure 6 by definition. We took the Reviewer's suggestion in the revised text:
"We find that although peak values of AR probability and IVT amount are sustained for two consecutive days, dry-bulb temperature decreases on the second day of the pair due to the shifting position of the ridge-trough system, while specific humidity remains nearly as high as on the first day."

L221: *Here, do the AR probabilities stand for the relative risk? I would refer to relative risk when corresponding, consider using the same therminology used in the Methods Section 2.4 to avoid confusion.*

We have changed the wording throughout the manuscript to 'relative risk' to ensure there is no confusion between this and other metrics.

L221&223: *First is used "hatched" to make reference to the tale colour information in subplots a,d,g and later is described as "contours". I believe it refers to the same thing, I would use the same wording to avoid confusion.*

The hatching is indeed only within the contours, and so we have changed both instances to 'hatched contours' for clarity.

4. Technical correction
When referencing figures, the authors used 2 different forms, abbreviation as "Fig. ##" when added directly to the text or without abbreviation like "Figure ##" when added in brackets, I would be consistent and only use one form in the manuscript.

These cases have all been standardized to "Figure #".

Figures 1 and 3 *do not present any text in the colorbar (like "Relative risk of AR/humid-heat interaction"). I would be clear in the plots what the colour values stand for, not only in the caption of the Figures.*

We have added a colorbar label (reading "Relative Risk") in Figure 1. In Figure 3, we enlarged the colorbar label because it was quite small and easy to miss:

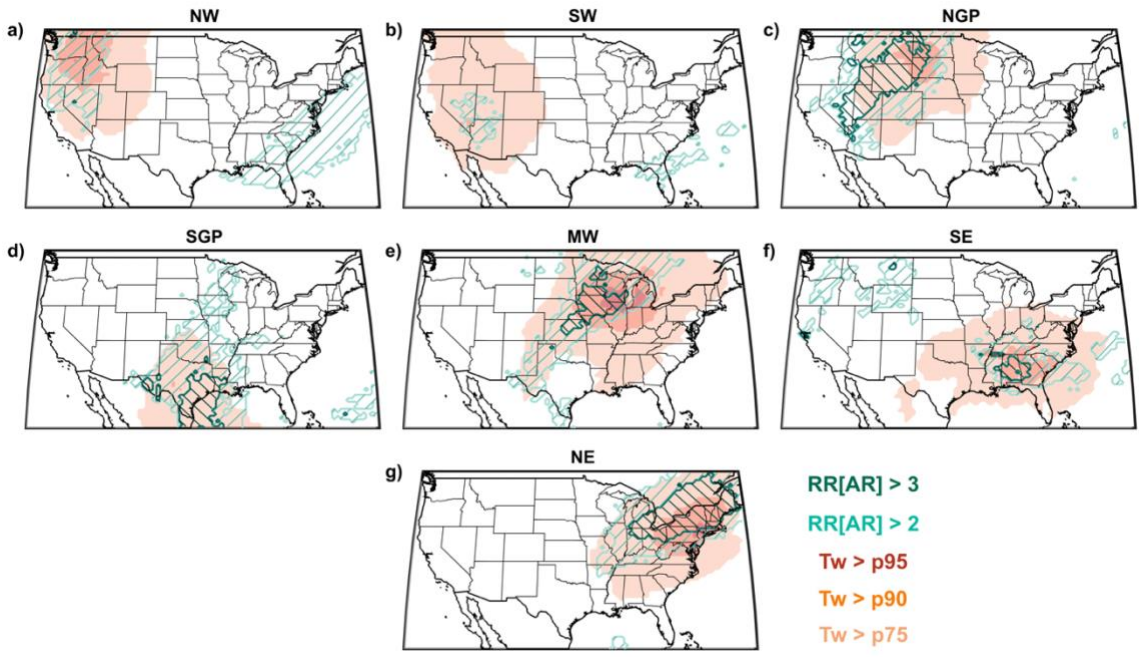


Figure 2: *Both y-axis for subfigures a) and b) have the same label, but stand for two different relative risks (the second is controlled by z500) as described in the caption. Consider using different y-axis labels for clarity.*

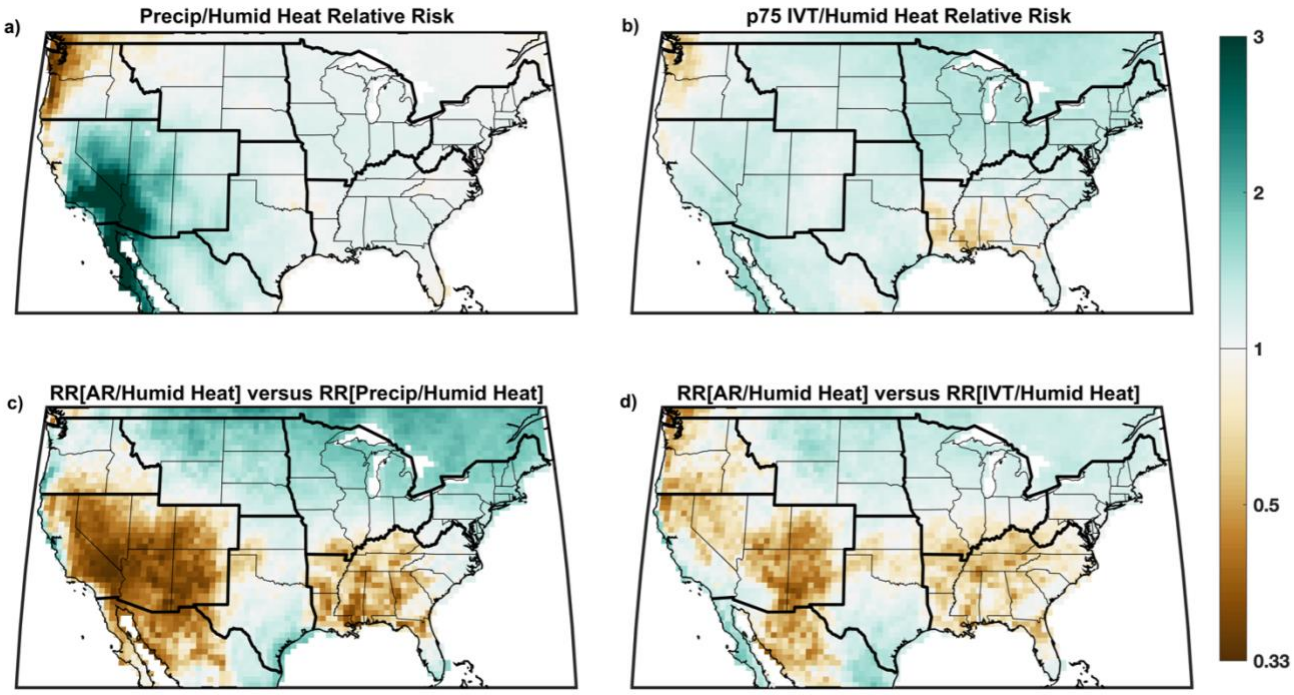
We have changed the y-axis label of panel (b) of Figure 2 to "Z500-Normalized Relative Risk of Extreme Tw".

Figures 3 and 5: *In the green labels are written "Pr[AR]" and in Figure 5 the blue labels are written "Pr[P]". I understand they stand for relative risk, but I would rather rewrite them as RR[AR] and RR[P]. Pr[] can be confusing as it is used in other places with a different meaning.*

Thank you for the suggestion. The figure labels have been corrected in this way for Figures 3 and 5, as well as for Figures A2-A7.

Figure 4: *Consider putting labels on top of each subplot for clarity, as it is done for Figure 5.*

We have added subplot labels to Figure 4:



L221&223 (Figure 5 caption): *The caption of subplots (b,e,h) and (c,f,i) are a bit confusing. The information in the caption does not always correspond to what is written in the legend at each subplot. Also, you write "as in (a,d,g)", but instead of showing relative likelihood, you show percentiles. I would recommend making this caption description clearer not to confuse the reader.*

We appreciate the catch — the (c,f,i) subplot labels referred to a slightly outdated figure version. We have revised them and added new text to be clearer without excessively lengthening the caption. The revised portion of the Figure 5 caption is:
"(c,f,i) As in (a,d,g) but for precipitation [P] and integrated vapour transport [IVT], with intervals for the former representing a relative risk of 2, 1.75, and 1.5 on composited humid-heat days, and for the latter the 80th, 75th, and 67th percentiles. These specific thresholds were chosen for visual clarity."