Reviewer 1:

This paper discusses making use of copulas and extreme value distributions conditional on different storm types to quantify annual exceedance probabilities of joint events at multiple locations, with a focus on two case studies in the USA. Events are described as being driven by non-tidal residuals or by rainfall caused by, in part, tropical and extratropical cyclones. The paper looks to analyse the relationships between flooding impact and these different drivers based on meteorological observations (ERA5) and hydrological modelling (UTide). Extreme events were identified using a Peaks-over-Threshold approach with a simple Independence/declustering criterion, and attributed to cyclones with a simple spatio-temporal proximity metric. Joint AEPs are based on "AND" scenarios (occurrence of events at all locations at the same time) and use a variety of copula families to determine these.

Overall, looking at the two similar problems of event driver attribution and simultaneous events is fairly novel. Although the constituent parts are known standards, the combination is not one this reviewer has come across before, which is much to the paper's benefit. At times, this paper is spinning a lot of plates at once, but ultimately the conclusions are clearly shown by the methods and data. The stated flexibility to more than two sites simultaneously is mentioned but not demonstrated.

Overall, this is a strong paper, which needs some adjustments to improve the communication of the methods and the conclusions.

The authors would like to thank the Reviewer for providing thoughtful comments. Below, the Reviewer can find our responses to each comment, including how we will address each of them in the revised manuscript.

Major issues:

The POT event declustering is quite simplistic and doesn't count for the varying speed of response of the flow to rainfall at each location.

Thank you for the comment. We agree that the varying speed of the runoff affects flood depths in the catchment. However, the objective of this paper is to assess the compound flood potential, in other words the probability of different combinations of rainfall and coastal water levels to co-occur. The physical response of the catchments to these compound rainfall and NTR events is therefore out of the scope of this study, and it will need to be assessed using numerical flood models that are able to represent many characteristics of the catchments that can affect resulting flooding (such as elevation, slope, soil type for infiltration and friction). The output of the framework we propose can act as boundary conditions for these numerical flood models.

The reference to "high" and "low" return periods from line 330 onwards needs to be more specific. The exact choice of cutoff is not important, but stating what it is and sticking to it is important.

The sentences will be changed as follows specifying the respective range of return periods within the brackets.

The sentence starting from L331:

“To quantify the relative contribution from each of them, we calculate the ratio of AEP contributed by TCs to the total AEP along the isolines (Fig. 10). In Gloucester City, more than 60% of the AEP of low probability events (i.e., events with return periods above 50 years) are associated with TC events while more than 70% of the AEP of high probability events (i.e., events with return periods below 20 years) are associated with non-TCs. In St. Petersburg, when both NTR and RF are extreme, TCs mainly drive the joint AEP. “

The sentence starting from L353:

“This becomes more noticeable for events with higher AEPs (i.e., events with return periods below 10 years), especially in St. Petersburg where it reaches up to 25% reduction. However, in Gloucester City, this impact becomes negligible for rare events (i.e., events with return periods above 50 years).”

The sentence starting from L434:

“Neglecting RF events that are likely locally generated can lead to an underestimation of the overall AEP (or overestimation of the return period), particularly for more frequent events (i.e., events with return periods below 10 years) (see panels (a) and (b) of Fig. 11).”

The sentence starting from L447:

“Overall, excluding non-classified events leads to smaller changes in the joint probabilities in Gloucester City compared to St. Petersburg, especially when focusing on rare events (i.e., events with return periods above 50 years) (see panels (e) and (f) of Fig. 11).”

Using high numbers of different copulas and distributions can make it hard to compare results, especially between sites. Using more general distributions (e.g. kappa) or more general copulas and sticking to one overall best choice may make it easier to interpret.

Our ultimate aim is the robust estimation of return levels which we believe is best achieved by selecting the best copula for each location with its particular dependence structure. As the results show, the correlation between the flood drivers can vary considerably across sites and between the TC and non-TC samples. Therefore, imposing a single copula family (which may not be a good fit in all cases) could mischaracterize the dependence structure. While it may make cross-site comparisons more straightforward, we would rather obtain the best results possible for each location. Then, when comparing the results of two sites, the derived probability distributions at each site are used, which are derived using the four best copulas for each stratified-conditional sample.

Minor issues:

The heavy use of abbreviations does make this harder to read, consider sometimes switching back to fully spelling out terms like rainfall.

We agree with the Reviewer that the overuse of abbreviations might make reading difficult. Therefore, we have removed those abbreviations that were used less than 25 times along the manuscript (e.g., Akaike information criterion (AIC) and general circulation models (GCMs)). All others are used extensively.

Check your references in the main body match those in the bibliography.

Thank you. The bibliography was re-checked and corrected.

Some paragraphs could be swapped for tables (paragraph around line 168, paragraph around line 330

We will revise the section at line 330 to define low/high return periods (RPs) as suggested by the Reviewer, and now we believe it is clearer in the paragraph format. Additionally, there is now more information that may not well fit into a tabular format.

For the section around line 168, we propose including a table that compares metrics for Gloucester City and St. Petersburg. This table would facilitate a direct comparison by listing the thresholds and number of peak-over threshold events. However, this addition may increase the length of the manuscript, as the table would need to be positioned above the corresponding text. Therefore, we kept the paragraph as in the original manuscript and hope this meets your approval.

line 252: numbered lists should be presented as such, with a new line for each for readability.

Agreed, the text has been changed as recommended by the Reviewer.

Figure 4: It is unclear what is exactly meant by "natural variability", and how it differs from the significance shown by the red circles.

In the manuscript, Figure 4 shows the changes in Kendall's τ between non-tidal residuals and rainfall derived from a 30-year moving time window. Each circle represents the calculated Kendall's τ for the peak-over threshold events during the respective 30-year period. The circles are marked in red when Kendall's τ is significant ($\alpha = 0.05$) for the respective 30-year period. We define the natural variability by randomly sampling 30 years many times (10,000 iterations were used) and calculating Kendall's τ. The range (gray band) corresponds to the 95th and 5th percentiles of the calculated Kendall's τ from the sampling process and it provides information about how τ could change by random chance over time due to natural variability. This is different from the individual τ values being significant or not.

We have added a sentence (in L279) in section 4.1 to clarify how the natural variability is defined and calculated.

"Fig. 4 shows the changes in Kendall's τ between NTR and RF derived from a 30-year moving time window and the range of the natural variability of the correlation. We define the natural variability by randomly sampling 30 years many times (10,000 iterations were used) and calculating Kendall's τ for each sample."

Figures 7 and 8 are very busy. Consider splitting into more figures or remove elements of the figures which do not directly contribute to your conclusions.

Thank you for the comment. For more clarity, the layout is changed to three rows and two columns (previously it was 2 by 3). The number of events in each panel and the name of the selected copula family were removed and will be added to the supplementary material (table).