Reviewer 2:

This manuscript by Chen et al., presents an analysis on compound flood hazard for the New York City area. The analysis is focused on compound events of precipitation and storm-surge that are driven by different storm events classified as tropical cyclones (TC), extra-tropical cyclones(ETC) and neither events. Results are also presented for "all" events considered, to highlight the differences in return period of the hazard when frequency analysis does not consider event type. Results shown suggest that despite the fact that the frequency of compound surge and rain events is low, the compound risk associated to TC events need to be assessed separately to avoid underestimation of the risk. Analysis has been based on a long record of hourly rain and tide gauges. Carrying the analysis at an hourly scale offers clear advantages, with respect to past works focusing on daily, on the identification of "simultaneous" rain-surge events and investigation of lag of the peaks (from rain and surge) overall.

Overall, the manuscript is clearly written, and the discussion and conclusions are supported by the results presented. Furthermore, the analysis at hourly scale and the event-type investigation offers novel elements for this type of work. Most of my major concerns on the methodological framework have been acknowledged by the authors themselves in section 5.4 "Limitations and simplifications", a fact that I appreciate because at the very least demonstrates that the authors understand and openly acknowledge the limitations of their approach and the complexity of the problem under study.

The authors would like to thank the Reviewer for their insightful comments. Below are our responses to each comment, including the planned adjustments to the manuscript.

Below I list some additional comments (mostly minor) for the author's consideration.

1. Thinking of estimation of lag or equivalently identification of "simultaneous" extremes of rainfall and surge, and considering that timing for the two variables is derived from different locations in space (tide gauges for surge and rain gauges for rainfall), there is some potential effect therefore on lag estimation. The authors somewhat refer to this effect in lines 394-395 and mention that this is further discussed in Section 5.3, but it is not discussed any further in that section. I think that elaborating further on this (and potential implications on the findings or methodology overall) is required.

We thank the reviewer for highlighting this important consideration. We acknowledge that using tide gauge data for surge and rain gauge data for rainfall can introduce spatial mismatches, potentially affecting the precise estimation of lag times between the two flood drivers. In our revised manuscript, we will expand Section 5.3 by adding:

> Different spatial locations of gauges may lead to uncertainties in defining "simultaneous" extremes. However, the timing differences of NTR across New York Harbor, e.g. in Jamaica Bay, off Manhattan, or in Newark Bay, is at most 30 minutes based on shallow water wave travel time (similar to tide) from offshore to reach these locations which have pathways with differences of at most 20 km. For single-gage analysis of rainfall-surge timing, these location differences may help explain different rank correlations. but for the joint probability analysis and lag time histograms we are using a spatial average rainfall, which captures regions surrounding the tide gauges well and should introduce very little timing difference.

2. Line 46: "and frameworks". Elaborate on what frameworks you refer to here, it is current statement is quite vague.

We appreciate the comment regarding the lack of clarity. In the revised version, we will specify that by "frameworks" we refer to established multivariate statistical and probabilistic modeling frameworks—such as Copula-based models and joint return period analyses—that have been widely used to assess the dependency between compound flood drivers. This clarification will be added to enhance the reader's understanding.

3. "Metro-scale rainfall". The reason for estimating average over that scale and its incorporation in the overall analysis is not clearly explained.

We agree that the role of metro-scale rainfall requires further explanation. In our revised manuscript, we will clarify that calculating a spatial average over rain gauges within a 30-km radius helps smooth out the localized variability in rainfall, giving a perspective more reflecting an integrated hourly (or longer) effect on flooding. Averaging gauges over such a spatial scale is a common approach for compound flood studies. As already noted in the text, we analyzed both the sewershed scale rain data (single-gauge) and the metro-scale rainfall in our analyses, for different perspectives.

4.L150: "we eliminate peaks that occur within 5-day windows". I assume that you mean that only the max peak within a 5-day window was retained(?). Please clarify. 5 days is admittedly a long duration for small scale pluvial flood events.

We confirm that our methodology retains only the maximum peak within any 5-day window to ensure event independence. This 5-day window was chosen to account for the typical maximum duration of cyclonic storm events, but can also avoid double-counting small-scale pluvial events that occur from the same weather system. We will clarify this point in the revised manuscript and discuss the rationale behind choosing a 5-day interval.

5. What is the total (and per class) sample size of rainfall and NTR peaks? The exact numbers should be reported for the readers to appreciate the sample size involved in this analysis.

We agree that providing explicit numbers will help readers better appreciate the robustness of our analysis. In the revised manuscript, we will include a table that summarizes the total number of rainfall and NTR peaks extracted from the 75-year record, along with a breakdown by storm type (TC, ETC, and non-cyclonic/convective events).

6. Figure 3 is not visually appealing (just the opinion of this reviewer), I wonder if you could improve how you convey the information in this figure.

We appreciate the reviewer's aesthetic feedback. We will revise Figure 3 by exploring alternative color palettes (for example, those recommended by ColorBrewer) and refining the graphical layout to enhance clarity and readability. Our goal is to improve the figure's ability to convey key information regarding the annual frequency of compound events across different storm types.

7. Figure 6: Have you accounted for the bin width when you calculated the values? If not the y-axis should be labeled frequency instead of probability density.

We agree with the reviewer that the y-axis for Figure 6 should simply be "Frequency (%)" and we will change the figure.

8. Finally, considering that the compound flood hazard (i.e. total flood depth resulting from rainfall and surge) is not explicitly considered and thus conclusions on the influence of the storm types on the hazard cannot be directly derived without coupled simulations (as the author acknowledge), I would recommend modifying the title to "Influence of Storm Type on Compound Flood Drivers" or something along those lines.

We appreciate the suggestion. In recognition that our analysis primarily addresses the influence of storm type on the flood drivers (rainfall and surge) rather than on the combined flood depth, we agree that a revised title would more accurately reflect our study's scope. Accordingly, we agree with changing the title to "Influence of Storm Type on Compound Flood Drivers." We believe this modification will better set the expectations regarding the focus of our work.