Reviewer 1:

The article "Influence of Storm Type on Compound Flood Hazard of a Mid-Latitude Coastal-Urban Environment" explores the differences between tropical cyclone (TC), extra-tropical cyclone (ETC) and non-cyclone (NC) as drivers of compound flooding to New York city, USA. The authors examine historic time-series of hourly rain and tide gauge data, using dependence and joint probability analysis methods, to explore the potential influence of storm type on near-simultaneous pluvial and storm surge flooding events. The study found that TCs dominate the most extreme pluvial/storm surge compound flood events, but ETCs are responsible for the majority of moderate and high frequency occurrences. There are important magnitude and lag differences depending on coastal location.

Strengths:

1. **Aim and objectives:** the aims and objectives of the study are clearly stated. The use of long time-series of gauged data is to be welcomed.

2. **Discussion and conclusions:** the focus on compound floods due to TC- and ETC-linked drivers is timely, given the rising frequency of extreme weather even within temperate zones, due to climate change.

3. **Statement of limitations:** it benefits this study that simplifications and assumptions are clearly stated. This text provides context and is a source of ideas for future research.

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.

Minor issues:

1. *Sea level rise trend: it was not completely clear how the trend of sea level rise has been removed from the 75-year time-series data measured at the tide gauges. Could the authors please expand on this?*

   We appreciate the reviewer's comment. We will clarify this methodology in Section 3.2 and include additional details to ensure reproducibility.

2. ***Distinction between TCs, ETCs and NCs/convective storms:*** *It was not clear to me how the authors categorized the different storm types in section 2. Was this pre-assigned to each storm by the National Hurricane Centre, or was a threshold (e.g. as defined by the Saffir-Simpson scale) applied afterwards? This is key information for anyone wishing to reproduce the study.*

   The categorization of storm types (Tropical Cyclones (TCs), Extratropical Cyclones (ETCs), and "Neither") was pre-assigned using records from the National Hurricane Center (NHC) and the reanalysis data (Section 2.2.3). Storms in the HURDAT dataset were considered TCs, given that they are nearly always either TC or post-tropical cyclones. It is very unusual for an extratropical cyclone to transition to tropical before passing NYB, with one case being the 1991 perfect storm.

   We will expand Section 3.4 to explain this classification process in greater detail.

3. ***POT approach:*** *could the authors please expand on the selection of the top-5 ranked rain/surge events each year? For context, it would be interesting to know how many of these events (out of all 75 x 5 events picked, per gauge), were categorized TCs, ETCs or NC/convective storms. A simple table would be enough. Only being able to capture a few TC events in the record, even with a long time-series, has to be recognized as an unavoidable limitation.*

   We agree with the reviewer. First, we already have Figure 3 to describe the annual frequency of each storm type within the top-ranked events for the Battery station. The selection of the top-ranked rain/surge events per year was based on their magnitude using the Peaks-Over-Threshold (POT) approach, representing an average annual exceedance frequency of 5/year (not the same as "top-5" per year).

   Out of the total 75 × 5 events across the 75-year dataset per gauge, we will add a new table summarizing the breakdown of TCs, ETCs, and Neither for both the Battery and the Kings point stations in the supplementary material.

   This table will highlight the relatively low frequency of TCs due to their rarity. In addition, we will add more text in the limitation section of the paper to acknowledge the challenge of capturing these events even in a long time series.

4. **Return period assumptions:** *the data are not longer than 75-years, how do the authors defend the calculation of return periods in excess of this (e.g. in figure 8)?*

Thank you for noting this point.

Due to the importance of extreme events (e.g. 100-year return levels) in planning and insurance, using shorter durations (e.g. our 75 years -1948 to 2022) of observed data to predict return periods in longer periods is common for univariate (e.g., Arns et al., 2015) and bivariate (e.g., Zscheischler et al. 2017) analysis.

In Table S1, the p values demonstrate that the copula models and thus return periods are plausible. Also, the scatterplots in Figure 8 also qualitatively back up our conclusion that TCs lead to worse joint hazards than other storm types. Importantly, we are not citing the exact values of rain and surge pertaining to specific return periods, we are only showing that an analysis of "ALL" data leads to smaller joint 50- to 200-year return levels for both flood drivers than analysis of TCs only. Therefore, our conclusions are robust to uncertainties in the fitted copula models.

5. **Assumption of stationary storm surge over time:** *While it is stated that this first baseline assessment does simplify conditions, it could also be worth mentioning that recent research has identified that elsewhere, storm surge extremes are not in fact stationary, over similar time-scales (e.g. Calafat et al 2022, DOI: https://doi.org/10.1038/s41586-022-04426-5)*

We agree this is worth mentioning and will add text mentioning the Calafat paper in the Discussion section on storm stationarity assumption.

6. **Figures:** *the majority of figures would be difficult to read for those who print in B&W, or are color-blind. Would recommend a different color palette (colorbrewer2.org, for example, suggests great color combinations that overcome this problem). In addition, would suggest that Figure 1 would benefit from (a) a simpler (line drawing) background rather than satellite imagery; (b) an inset, or wider view, to illustrate the NYC location with more of the Long Island Sound and Atlantic visible (to better understand storm surge, and significance of storm orientation at each gauge location); and (c) to perhaps reconsider the color scheme of gauge location points for the reasons stated to above. Additionally, figure 8 would benefit from larger font in the x-, and y- axis labels.*

We appreciate the feedback on figure accessibility and will address these concerns:

(a) We will simplify the background of Figure 1 to a line drawing for clarity.

(b) We will add an inset or wider view in Figure 1 to provide context for the NYC location and surrounding regions.

(c) We will revise the color scheme for all figures to use colorblind-friendly palettes (e.g., from ColorBrewer2.org).

(d) We will increase the font size of axis labels in Figure 8 for readability.

7. **Statement of relevance:** *the manuscript might benefit from a clearer description of the significance of the results of this study, which focuses on a relatively small urban watershed referencing a small number of gauges, to the current scientific knowledge of pluvial/coastal compound flooding. How do these findings contribute to the scientific conversation?*

We agree that the primary contributions were not coming through clearly. To improve this, we will enhance the start of the Discussion section to better articulate the broader implications of our findings. The first paragraph already presents the value of our separation of TCs and other storm types, which has rarely been done for compound hazard research. The second paragraph already presents the value of using hourly data and the new knowledge that is obtained which would not be possible using past common methods of daily data and loose definitions of overlap. For both paragraphs, we will add topic sentences that make clear how these two core approaches of our paper contribute to the broader field.

Overall:

Because of the use of hourly time-series data, this study provides useful insights into how lag time, magnitude, and orientation of storm-linked drivers all contribute to the state of flooding within an urban watershed of high economic value. The use of this more discrete data, creates a useful distinction between impacts in compound flooding due to TCs, ETCs, and convective storms. The study would benefit from clarifying some details of the methodology and results, as detailed above.

Technical corrections:

- *L486 "in toto"?*

  Thank you! It is a typo. We will replace it with "in total" to improve readability.

- *L114 - how long is the data collected at Battery gauge?*

  The text will be changed to clarify that the Battery gauge has near-complete long-term temporal coverage during the period of hourly rain data from 1948 to 2022, spanning approximately 75 years.

- *L 143 – what is a "sewershed"?*

  A "sewershed" refers to an area of land where all surface water drains into a common sewer system, similar to a watershed but specifically for urban stormwater and wastewater systems.

- *L165-L167. At a single gauge is this statement correct? This feature of storm surge is known due to onshore and offshore winds in different quadrants of the TC position; however usually one tide gauge records rising levels due to onshore winds, and a neighbor some km away would (hopefully be well-placed to) capture the negative surge due to offshore winds. Of course this effect changes with cyclone path/coastline orientation, and cyclone size.*

  We agree that the text in this paragraph was confusing, and felt that it was not needed. We will eliminate the statement. However, we note that there are no known historical cases at NYH where there are such large differences in surge over small distances of a few km. Surge is typically very similar across NY Harbor, though it can be different at Long Island Sound (Kings Point), which was the reason we separated the analysis into these two areas.