Reviewer 2:

This paper is well-written and scientifically sound. I would like to see more discussion on the following topics:

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 as per the NHESS review process.

[1] you end up with results for RF and NTR. Please recommend how to combine these results in a flood risk analysis

The main objective of the paper is to provide a framework that can be used to assess the compound flood potential and to derive joint probabilities of flood drivers accounting for the different statistical characteristics depending on the storm type that generates them. Procedures for combining the return level estimates among different populations in a univariate analysis are available (Barth et al., 2019), however in the bivariate space it is an active area of research with most similar studies focusing on a single population of events (e.g., TCs in Kim et al., 2023).

Flood risks can be estimated using different approaches/definitions. If we assume the IPCC definition, flood risk is a combination of flood hazard, with exposure and vulnerability. Our results can be used to estimate the flood hazard. Again, there are several approaches that can be followed. As a common approach, combinations of RF and NTR of specific probabilities (e.g. 100 years) can be selected to derive the resulting flooding from these events using a flood model and thus estimating the flood hazard. In another approach, the statistical analysis is performed on some response variable (e.g., flood depth at a given location) after running a large number of events comprised of synthetic RF and NTR through a flood model. The authors are currently working on a framework to produce time series of NTR and rainfall fields for these synthetic events, as they are required as boundary conditions by numerical flood models. To summarize the above process, the following part is added to the discussion section (where the limitations are discussed).

"To extend the proposed framework for fully characterizing the compound flood risk, the statistical approach can be combined with hydrodynamic numerical models (so called hybrid modeling; e.g., Moftakhari et al. 2019) to estimate flood inundation. However, analyzing only the most likely event (even though it may be the most plausible given the observations) does not capture the range of flood levels that could be generated by different combinations of flood drivers (NTR and RF) along an isoline. One way to address this limitation is to sample an ensemble of events (peak NTR-RF combinations) along the isoline and run them through flood models. Alternatively, a response-based approach can be employed, which involves simulating flood hazard for a large number of synthetic events from the multivariate statical model and then performing the statistical analysis on the response variable of interest (e.g., flood depth at a given location). The latter is computationally demanding, possibly necessitating the use of a surrogate model, however the return level estimates are likely to be more robust than when adopting an event-based approach (Jane et al., 2022). The simulated probabilistic flood depths and extents can then be incorporated with exposure and vulnerability data to perform a comprehensive flood risk assessment."

[2] the isolines method has some shortcomings in my view. The paper hints that these can be used to define a single event (the one on the isoline with the highest probability density), However, in most

compound flood risk analysis there is no single "most representative event". In coastal zone there are locations close to the coastline for which the main flood driver is the peak sea level (surge) and there are locations more inland for which the rainfall is the dominant flood river. For the former, an event with extreme NTR and moderate RF is the most relevant, for the latter an event with extreme RF and moderate NTR is the most relevant.

Thank you for the very useful comment. In this paper we adopt the "most-likely event" strategy, introduced by Salvadori et al. (2011) and subsequently used in other studies (e.g., Jane et al., 2020), as a straightforward method to derive possible design events associated with a given return period. Analyzing just the most likely event (even though it may be the most plausible given the observations) will not capture the different flood levels that could be generated through the different combinations of flood drivers (NTR and RF) along an isoline. One option to overcome this limitation is to stick with the event-based approach and to sample an ensemble of events along the isoline, and then run them through flood models. Another option is to use a response-based approach, which involves simulating a large number of synthetic events from the multivariate statical model, run them through a flood model, and then calculate the empirical return water level at a given location using the water levels generated by these events. The latter is computationally demanding, possibly necessitating the use of a surrogate model, however the return level estimates are likely to be more robust than when adopting an event-based approach. In response to the Reviewer's comment, the last paragraph of Section 3.5 will be revised as follows:

“Although any combinations of NTR and RF along a given joint probability isoline have the same return probability, most hydrology-related engineering design approaches still rely on a single design event. Therefore, the “most likely event” strategy, introduced by Salvadori et al. (2011) and utilized in subsequent studies (e.g., Jane et al., 2020), is employed here. To quantify the relative probabilities of events along specific quantile isolines, we obtain 10^6 combinations of NTR and RF by sampling from the fitted copulas, ensuring that the relative proportion of extremes is consistent with the empirical distribution. Then the relative probability along the isolines is calculated by the kernel density function of the simulated sample. The location of the “most likely” event is assigned to the point with the highest relative probability density on an isoline (Salvadori & Michele, 2013).“

The following part is also added to the discussion section that clarifies the reviewer’s concern. (Where the limitations are discussed).

“However, analyzing only the most likely event (even though it may be the most plausible given the observations) does not capture the range of flood levels that could be generated by different combinations of flood drivers (NTR and RF) along an isoline. One way to address this limitation is to sample an ensemble of events (peak NTR-RF combinations) along the isoline and run them through flood models. Alternatively, a response-based approach can be employed, which involves simulating flood hazard for a large number of synthetic events from the multivariate statical model and then performing the statistical analysis on the response variable of interest (e.g., flood depth at a given location). The latter is computationally demanding, possibly necessitating the use of a surrogate model, however the return level estimates are likely to be more robust than when adopting an event-based approach (Jane et al., 2022). The simulated probabilistic flood depths and extents can then be incorporated with exposure and vulnerability data to perform a comprehensive flood risk assessment.

furthermore some minor comments in the attached document

line 74: Not sure I agree, see for example: https://link.springer.com/article/10.1007/s11069-024-06552-x

Thank you for pointing out the reference. It was published after we submitted our manuscript and hence couldn't include it. In our paper we claim that a "a comprehensive multivariate statistical framework for assessing compound flood potential" does not exist. The study that the Reviewer mentioned simulates the flood depths using physics-based models and applies univariate extreme value analysis to the simulated flood depths and combined yearly exceedance frequency of TCs, and ETCs following Dullaart et al. (2021). Therefore, this study does not really fall into the category of multivariate statistical frameworks assessing compound flood potential.

However, it is still relevant to the scope of our paper, and we now cite it in L46. An additional part was added to describe their work on L69 as follows:

"Nederhoff et al. (2024) addressed this aspect by employing a compound flood model for the coast of the United States, from Virginia to Florida. They separately simulated the total water levels induced by TCs and ETCs to assess their relative contributions and followed the approach outlined by Dullaart et al. (2021) to calculate the combined return water levels."

Line 141: Isn't this essentially the same as applying an areal reduction factor (ARF)?

Yes, the concept is similar to "areal reduction factor (ARF)". The ARF is generally applied to point rainfall estimates whenever an area (around the rain gauge) is large enough for rainfall not to be uniform. However, the Philadelphia International Airport rain gauge is not within the selected catchment (about 6 miles away from the selected catchment); thus, it is necessary to correct the bias due to the different locations (if any), additionally to what is corrected by ARF. Therefore, a simple method (quantile mapping) is used to transform rainfall gauge data to catchment representative quantities, and the term "bias-correction" is used to avoid any confusion.

Line 245: Why do you use sampling? You have the mathematical description of the copula and the marginals, so probability densities can be computed without sampling. Did you repeat this procedure multiple times with a different seed to test the variability in the results?

The authors agree that probability densities along the isoline can be calculated using the fitted copulas and marginals without sampling. However, this approach might not be straightforward since there are four copulas involved (conditioned on rainfall and NTR for both TC and non-TC) when estimating probability densities for the combined populations. Additionally, the same weight cannot be given to the relative joint return probabilities (along an isoline) derived from TC and non-TC copulas.

Therefore, a simpler method is employed here: sampling a sufficiently large number of NTR-rainfall combinations from each fitted copula to ensure that the relative proportion of extremes is consistent with the number of threshold exceedances in each corresponding sample. Then, the kernel density estimate is applied along the isoline for all the simulated NTR-rainfall combinations.

The sampling process was repeated multiple times, and it was noted that using 10,000 combinations is sufficient to obtain stable estimates for probability densities. However, following the Reviewer's comment and considering the low computation time for simulations, 1,000,000 simulations are now used to achieve even more stable estimates for probability densities along isolines. The text and Fig. 9 will be updated accordingly.

Line 313: To get more insight in the GOF for the highest values (extremes) a q-q plot could be valuable.

Thank you for the comment. The plots of parametric distributions with empirical distributions are used in the paper considering the importance of showing the confidence intervals. As per the Reviewers comment the quantile plots for each fitted distribution are shown below.

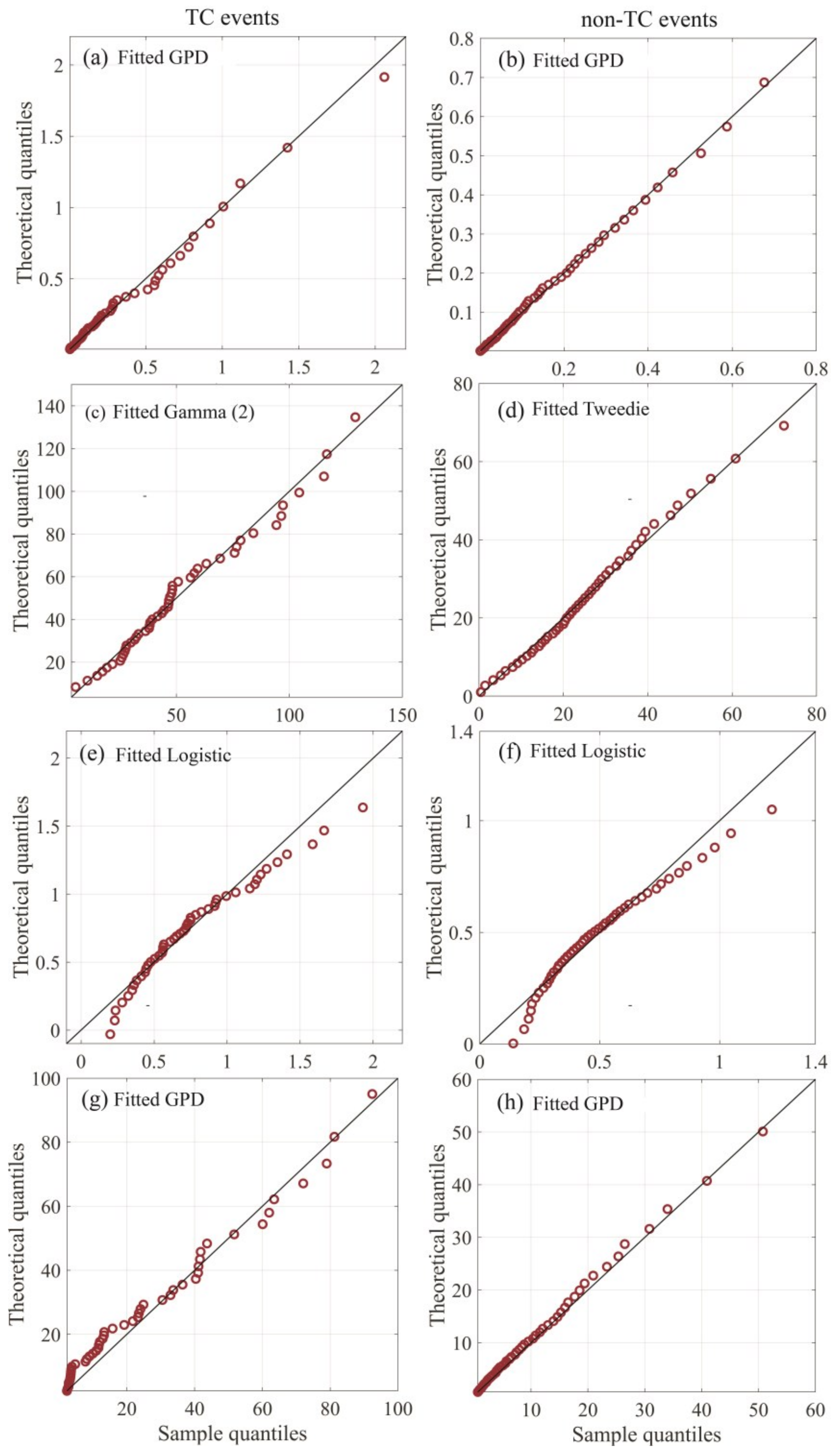

Figure R1: Q-Q plots for selected parametric distributions compared to the empirical distributions of Gloucester City, for TC events (left) and non-TC events (right). (a), (b) NTR POT events when conditioned on NTR. (c), (d) Maximum RF events corresponding to NTR POT events when conditioned on NTR. (e), (f) Maximum NTR events corresponding to RF POT events when conditioned on RF; (g), (h) RF POT events when conditioned on RF. The black line indicates the x=y line in each panel.

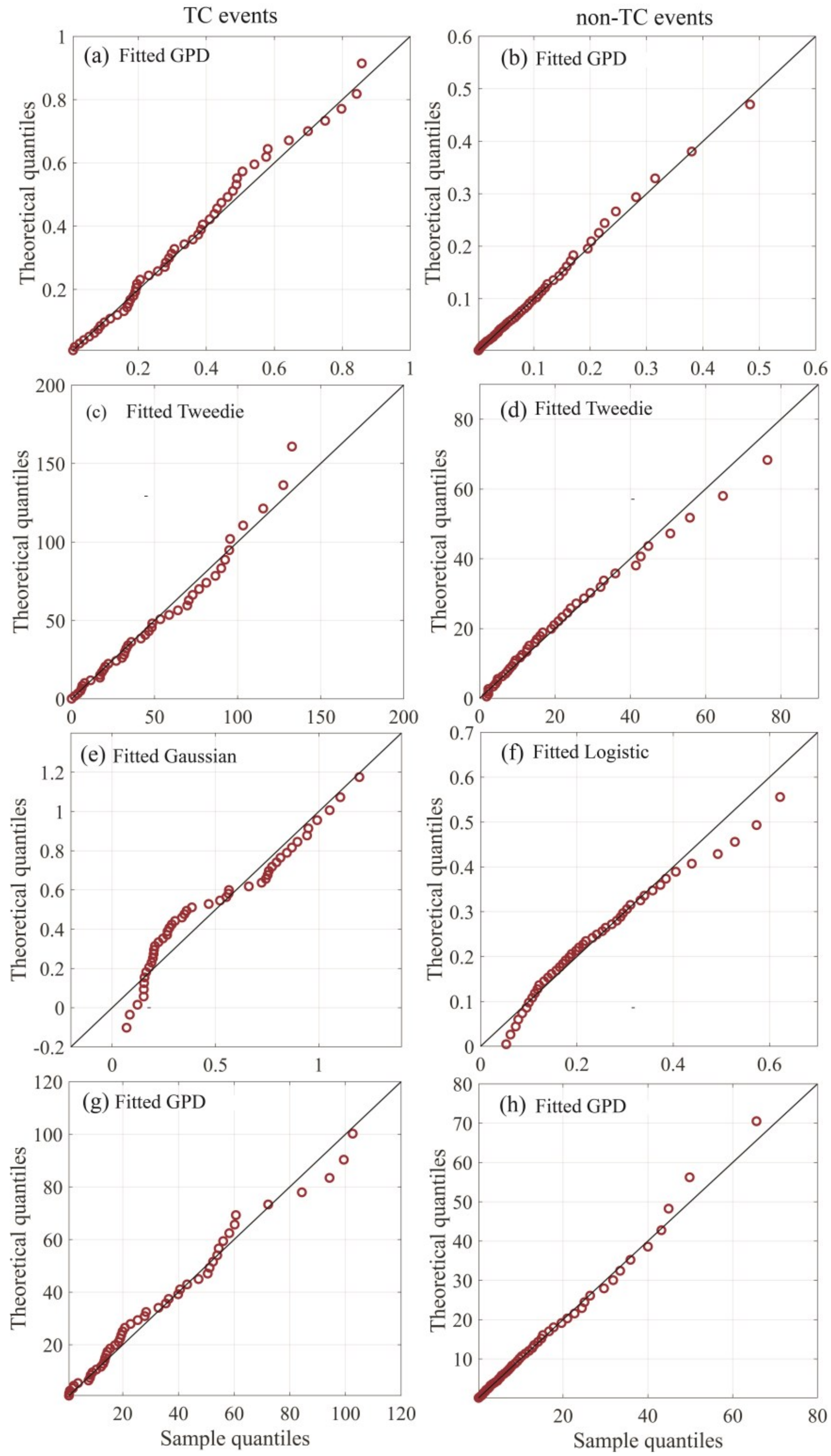

Figure R2: same as Figure R1, but for St. Petersburg