**Reply to reviewer's comments**

We thank the reviewer for the thorough reading of the manuscript and the valuable remarks that helped us to improve the manuscript. In the following, the original reviewer comments are given in italic and all line numbers and figure numbers refer to the submitted version.

**Reply to review of reviewer 2**

*As reviewer 2 in the last round of the review process, I would like to congratulate the authors on the revised manuscript. In my opinion adding a separation criterion to the de-clustering procedure has increased the robustness of the method, while paragraph structure and quality of the Figures are much improved on the initial submission.*

We thank reviewer 2 for the overall positive evaluation of our revised manuscript and very detailed comments.

*In my view, there are several issues that need to be addressed before the manuscript is ready for publication. Overall, the methods and analysis are robust, however the purpose of the work/ utility of the results remains unclear, with only a single possible useful application alluded to at the very end of the manuscript. The explanation of the statistics in the introduction requires attention and the discussion of existing studies could be revised to provide more insight.*

We address these concerns in the following sections.

*The authors really need to figure out the overall aim/purpose of the work and state it before the three tasks are introduced in the abstract. A reader needs to know why it is worth their time reading the paper. In the current text the aim of the work is too vague. For example, "It is important to" is subjective, and what constitutes a underlying mechanisms is not clear. "Our study focuses on the analysis of potential compound flood events" I could guess this from the previous sentences! "with the following contributions" they are steps rather than contributions.*

We made the following changes to the abstract starting from line 2:

Compared to the occurrence of single extreme events, co-occurring or compound extremes may substantially increase risks. To adequately address such risks, improving our understanding of compound flood events in Europe is necessary and requires reliable estimates of their probability of occurrence together with potential future changes. In this study compound flood events in northern and central Europe were studied using a Monte-Carlo based approach that avoids the use copulas. […]

Furthermore, we made the following addition to line 407:

This leads to the question of how the frequency of compound flood events might change for the various parts of Europe, which is vital for regional coastal adaptation.

*The first paragraph in the introduction zeros in on coastal flooding, however later in the introduction studies concerning "compound inland floods" are discussed. Due to the depth of the literature on compound flooding, I suggest omitting studies that do not involve discharge-*

*surge compounding.*

We removed the three references in line 36 in favour of a study that involves discharge-surge:

Several studies conducted over the last years have shown the importance and catastrophic nature of compound flood events for various locations.

One example is the flooding of Jacksonville (Florida) where the surge caused by the strong winds of Hurricane Irma stalled the fluvial discharge (Juarez et al., 2022).

*Phases such as "many studies" are used rather a lot in the introduction. The introduction should be crafted to describe the relevant findings of the previous studies in a way that engages the reader. Lines 60-63 are a good example of where the key findings of the individual studies should be stated rather than summarized with a semi-relevant generic statement two sentences later. Furthermore, it is not true that results cannot be compared, its just that differences may at least partially be down to the chosen approaches. The paragraph starting on L64 is a well-crafted thoughtful paragraph.*

We reduced the number of listed citations and further elaborated on the remaining ones.

Line 39:

Moreover, several studies have been conducted on a larger spatial scale in Europe. Considering data from 1901-2014 and gauges from northwestern Europe Ganguli and Merz (2019) found opposing trends in the magnitude of compound flood events depending on the latitude of the gauge. They reported increases at midlatitudes (47°N to 60°N) and decreases for gauges at high latitude (>60°N). Svensson and Jones (2002) analysed the dependence of high sea surge, river flow, and precipitation in the UK. They found a higher number of compound flood events on the western than on the eastern cost, while Paprotny et al. (2020) demonstrated that hydrodynamic models are capable of identifying real world compound flood events in north-western Europe.

Line 49:

Additionally, there have been studies modelling compound flood events in rivers on a local scale such as for the Zengwen River basin in Taiwan by Chen and Liu (2014), the Shoalhaven River in Australia by Kumbier et al. (2018), and the Fuzhou in China by Lian et al. (2013).

Line 60:

Nonetheless, there have been some studies that investigated the sensitivity of their results. Zheng et al. (2014) compared three classes of statistical methods and found that the point process method overestimated the dependence of extremes while the conditional method underestimated it. In a similar vein, Jane et al. (2022) assessed that their estimates of the potential for compound events were highly sensitive to the statistical model setup.

Removed sentence at the end of line 61.

Additionally, line 53 was modified:

A direct comparison of the results from different studies is hampered by the use of different approaches, data, analysis periods, and other factors.

*Most of the statements relating to the statistical modeling can be made clearer and some are erroneous. "They were introduced in Sklar (1959)" is an example of the latter. I suggest consulting with a statistician familiar with this field to tighten up the text regarding the copula modeling. A definition of a copula is also missing.*

We tried to present the arguments with more rigour and references to existing studies. The revised paragraph reads as follows:

Many studies utilised multivariate extreme value theory and copulas to describe the data distribution of two or more time series and investigate the dependence between extreme events (Hao et al., 2018). In climate research, the amount of available data points is often very small, with many studies operating at merely 30 extreme events. This can cause large uncertainties when trying to evaluate the tail dependence of the multivariate distributions (Serinaldi, 2013; Serinaldi et al., 2015; Joe, 2014). An alternative approach is based on Monte-Carlo based simulations where the dependence between joint extremes is studied by randomly rearranging one of the time series. Given our small sample size, in the following we used such an approach to avoid the uncertainties associated with the use of copulas in small samples.

* *L18: "Drivers for floods are storm surges, waves, tides, precipitation, and high river discharge" Perhaps worth noting that stretched of river where these drivers combine to exacerbate flooding are referred to as transition zones (e.g., Bilskie and Hagen 2018).*

We added this information to line 19:

The area of the river in which two or more of these drivers influence the water level are called flood transition zones (Bilskie and Hagen, 2018).

* *L29: "Compound flood events occur when large run-off from, e.g., heavy precipitation, leading to extreme river discharge, is combined with high sea level (storm surge)" there are many types of compound flood event, however I agree this is a good place to state that this is the type of compound flood the work will focus on.*

We added this information to line 29:

This study focuses on compound flood events that occur when large runoff from, e.g., heavy precipitation, leading to extreme river discharge, is combined with high sea level (storm surge).

* *L31: "In the following text we will note "potential compound flood events " as "compound flood events" for the sake of readability, with regards to literature see, e. g., Ganguli and Merz (2019), Jane et al. (2020), or Couasnon et al. (2020)." I suggest retaining this text!*

Added the sentence as suggested with an additional explanation as suggested in the first review.

Local flood protection and topography might prevent compounding extreme events from causing floods. Due to the size of the study area, we cannot take this into account and will denote these "potential compound flood events" as "compound flood events" in the following text for the sake of readability.

*     L33: "The occurrence of extreme flood events either simultaneously or in close succession can lead to severe damage, which greatly exceeds the damage those events would cause separately (de Ruiter et al., 2020; *Xu et al., 2022)." Most of the subsequent studies you describe only have one flood event caused by multiple drivers arising simultaneously.*

We changed the sentence in L33 to:

The occurrence of extreme flood and surge events either simultaneously or in close succession can lead to severe damage, which greatly exceeds the damage those events would cause separately (de Ruiter et al., 2020; Xu et al., 2022).

*     *L43: "All of them found that the assumption of independence between drivers leads to a huge underestimation of the occurrence rate of compound events." This is not true for all locations.*

Changed line 43 to:

Many studies found that the assumption of independence between drivers leads to an underestimation of the occurrence rate of compound events.

*     *L56-57: I believe these are all peaks-over-threshold approaches. There are also different peaks-over-threshold de-clustering methods e.g. storm window or runs method.*

We added the following information to line 118 since we feel that this is a better location for this information:

A critical element in the analysis is the definition of a de-clustering window such that subsequent events can be considered as independent. A frequently used window size is based for example on the typical duration of storms in the area (e.g. Harley, 2017; Camus et al., 2021). Here, we chose a de-clustering time of three days as used in other studies spanning larger domains (e.g., Bevacqua et al. (2019); Ward et al. (2018); Haigh et al. (2016)).

*     *L82: "Serinaldi et al. (2015) therefore concluded that those results are "highly questionable and should be carefully reconsidered". This is a very general statement, does he really say all copula models are highly questionable and should be carefully reconsidered.*

*The paragraph starting on L84 ends with "Consequently, we chose to study compound flood events by using a methodology that does not utilise copulas". It is strange because copulas are not mentioned at all in the rest of the paragraph. Copulas are not needed to calculate tail dependence.*

Please see our changes made in the "general comments" section since we rewrote this paragraph.

*     *L129: River discharge is traditionally denoted by a Q.*

We changed line 129 accordingly:

For the discharge of rivers we chose the 90th percentile $Q_{90}$ and for the sea level the 99th percentile $S_{99}$.

* L134: As I stated in my last review, I do not believe this is true. The trade-off between a large sample whilst ensuring the sample only contains actual extremes could be incorporated into the previous paragraph and then L134-137 could be deleted.*

We removed the sentence starting on line 134. We moved the sentences in lines 135-137 to the beginning of the paragraph.

Extreme events should be rare by definition, regardless of the river size, therefore only occurring scarcely throughout the year. This especially prevents the accidental analysis of events that are normally not considered as extreme. On the other hand, the choice of our threshold needed to take the limited data availability into account. Hence, we were forced to choose our thresholds low enough to ensure that enough points were available for robust statistical analysis.

* *L168: "This was similarly stated by Couasnon et al. (2020)." You can just cite Couasnon et al. (2020) at the end of the previous sentence.*

Moved the reference to the end of the previous sentence in line 168:

As a result, we would see a much lower number of compound flood events in the non-randomised data; therefore suggesting a false dependence (Couasnon et al., 2020).

* *L190-195: Please refer to my comment in the last round of reviews regarding the optimum location of river gauges for this type of study.*

Based on our reply to your comment we added the following information to line 197:

For our analysis, we utilised several model-based data sets which varied in forcing, regions and time frames.

The following sentence was added to line 198:

The simulated discharges are solely caused by the atmospheric forcing and the hydrological processes over land. The influence of the sea level on discharge in the estuaries of the rivers is not considered so that this influence (e.g., *Moftakhari et al. 2019*) does not cause problems in the determination of river floods.

* *L294 & 219: In English the abbreviation "vs." is typically shorthand for "verses". "v." is generally used for "version".*

Changed line 219:

Using E-OBS v. 22, HydroPy was driven by daily temperature and precipitation at 0.1° resolution from 1950–2019.

We assume that the reviewer meant line 204 instead of 294:

The HD model v. 5.0 (Hagemann and Ho-Hagemann, 2021) was set up over the European domain covering the land areas between -11° W to 69° E and 27° N to 72° N at a spatial resolution of 5 min (ca. 8-9 km).

* *L223: Grammar. This could work: "… found precipitation data from ERA5 to be of a higher quality than from EOBS."*

Changed line 223 as suggested:

Investigations by Rivoire et al. (2021) found precipitation data from ERA5 to be of higher quality than from E-OBS.

* *L266: Quite wordy. I would remove the reference to the eastern British coast in this sentence as it is discussed in the following sentence.*

Removed the reference in line 266 as suggested:

The east and south facing coasts of the Bothnian Bay and Bothnian Sea in the Baltic Sea, as well as Skagerrak, show the lowest frequencies of compound flood events.

* *L270: Doesn't this plot simply reflect the number of compound events in the record i.e., Figure 2. The west coast of France has a small number of compound events at most sites, therefore we'd expect it to be within 2 standard deviations. What about the other coasts?*

We agree that a very high number usually reflects being outside of 2 standard deviations. Nonetheless, the number of events depends to a certain extent also on the amount of data points. Long discharge events therefore offer a higher chance by randomness to coincide with a sea level extreme event. Analysing if the number of events is within 2 standard deviations is therefore a safeguard.

* *L271: Remove "or outside" it is superfluous.*

Removed "or outside" as suggested from line 271:

Utilising our randomisation method (cf. Sect. 2) yielded Fig. 3 that shows if the amount of observed compound flood events for each river is within the 2σ interval produced by the randomised data sets.

* *Figure 2 (caption): "The number of discharge and sea level extreme events was limited to two events per year." That is on average.*

Added to caption of Figure 2:

The number of discharge and sea level extreme events was limited to two events per year on average.

* *L280: The term "pattern" is ambiguous here.*

Changed line 280 to:

The overall pattern indicating that western coasts have the tendency of showing more events than expected by pure chance remains stable throughout these different data set combinations.

* *L283: Change "sections" to "periods".*

Changed "sections" to "periods" as suggested in line 283:

Next, we split the ECOSMO–coastDat3 and HD5–EOBS data into two 30 year periods, from 1960 to 1989 (Fig. 4b) and from 1990 to 2019 (Fig. 4c).

* *L321: Grammar. "flood compound events".*

Changed the order of words in line 321:

Anticyclonic Westerly is known to lead to precipitation in the area of the Baltic countries (Jaagus et al., 2010), which in combination with the south-eastern wind direction are responsible for around a third of the compound flood events in the Baltic and western facing Finnish area, due to the orientation of their coastline.

* *L342: Remove "such".*

Removed "such" from line 342:

Furthermore, the rivers were coloured red if the number of compound flood events is above the 2σ interval of randomised sea level data, blue if below the interval, and grey otherwise, as in Fig. 3.

* *L376: "In addition, we demonstrated that there exists a correlation between river catchment size and the number of compound flood events. It can be seen that, regardless of the estuary orientation, the number of compound flood events declined with increasing catchment size" All the information in the first sentence is contained within the next sentence. Also, they declined "on average", there is not an exact relationship.*

We combined both sentences and added the information that it declined "on average":

*In addition, we demonstrated that regardless of the estuary orientation, the number of compound flood events declined on average with increasing catchment size.*

* *L378: "The reason for this might be that rivers with smaller catchment areas are capable of reacting faster to precipitation that appears during the storm events, which also causes the storm surges." Did accounting for the catchment response times by including a lag weaken the trend?*

We thank the reviewer for the interesting suggestion. As stated in the publication text, we used a constant lag. We are aware that studies like Ganguli and Merz 2019 used a lag time that depended on the catchment size (formula 1 in their paper). Our concern is that especially for large rivers the lag would massively depend on the location of the precipitation. For example, the discharge of the Elbe would react much faster to precipitation in Hamburg than to precipitation near Prague. Therefore, it would be very challenging to quantify the lag for all the rivers in our study with a detailed investigation being beyond the scope of this work. Nevertheless, it is a very compelling topic that could be considered in future studies.

We therefore added to Line 395:
They could also attempt to quantify the lag for each catchment individually, which is currently challenging for large rivers since their lag depends on the location of the precipitation.

Reference:
Ganguli, P. and Merz, B.: Trends in compound flooding in northwestern Europe during 1901–2014, Geophysical Research Letters, 46, 10 810–10 820, https://doi.org/10.1029/2019GL084220, 2019.

*L388: Maybe change "for design events to test flood protection structures" to "such as design events used to assess the level of protection afforded by flood defence structures."*

Changed line 388 as suggested by the reviewer:

The lack of a parametric model impedes the possibility of deriving engineering quantities such as design events used to assess the level of protection afforded by flood defence structures.

*L389: "ensemble data" From climate models I assume. Another benefit to ensemble data is the higher spatial resolution. A drawback that it is numerically derived rather than observed.*

Changed line 389 to:

Future work can further examine these findings by using ensemble from climate models data over a longer time frame, e.g. 50 years and more.

Additionally, we added to line 393:

One potential drawback is the reliance on the capabilities of numerical models to adequately generate those compound extreme events.

------------------------

List of added references:

Bilskie, M. V., & Hagen, S. C. (2018). Defining flood zone transitions in low-gradient coastal regions. Geophysical Research Letters, 45, 2761– 2770. https://doi.org/10.1002/2018GL077524

Hao, Z., Singh, V. P., and Hao, F.: Compound extremes in hydroclimatology: a review, Water, 10, 718, https://doi.org/10.3390/w10060718, 2018.

Juarez, B., Stockton, S. A., Serafin, K. A., and Valle-Levinson, A.: Compound flooding in a subtropical estuary caused by Hurricane Irma 2017, Geophysical Research Letters, 49, e2022GL099 360, https://doi.org/10.1029/2022GL099360, 2022.

Moftakhari, H., Schubert, J. E., AghaKouchak, A., Matthew, R. A., and Sanders, B. F.: Linking statistical and hydrodynamic modeling for compound flood hazard assessment in tidal channels and estuaries, Advances in Water Resources, 128, 28–38,535 https://doi.org/10.1016/j.advwatres.2019.04.009, 2019.

Serinaldi, F.: An uncertain journey around the tails of multivariate hydrological distributions, Water Resources Research, 49, 6527–6547, https://doi.org/10.1002/wrcr.20531, 2013