Brief communication: The role of using precipitation or river discharge data when assessing global coastal compound flooding

Emanuele Bevacqua[1], Michalis I. Vousdoukas[2], Theodore G. Shepherd[1], and Mathieu Vrac[3]

[1]Department of Meteorology, University of Reading, Reading, United Kingdom
[2]European Commission, Joint Research Centre (JRC), Ispra, Italy.
[3]Laboratoire des Sciences du Climat et de l'Environnement, CNRS/IPSL, Gif-sur-Yvette, France.

Correspondence: Emanuele Bevacqua (e.bevacqua@reading.ac.uk)

Abstract. Interacting storm surges and high water-runoff can cause compound flooding (CF) in low-lying coasts and river estuaries. The large-scale CF hazard has been typically studied using proxies such as the concurrence of storm surge extremes either with precipitation or with river discharge extremes. Here the impact of the choice of such proxies is addressed employing state-of-the-art global datasets. Although being proxies of diverse physical mechanisms, we find that the two approaches show similar CF spatial patterns. ~~However, deviations~~ On average, deviations are smaller in regions where assessing the actual CF is more relevant, i.e. where the CF potential is high. Differences between the two assessments increase with the catchment size and our findings indicate that CF in long rivers (catchment $\gtrsim$ 5-10,000 Km^2) ~~is more accurately~~ should be analysed using river discharge data. The precipitation-based assessment allows for considering local rainfall-driven CF, and CF in small rivers not resolved by large-scale datasets.

1 Introduction

Compound flooding (CF) happens in low-lying coastal areas due to the interaction of high precipitation runoff and high sea level. The combination of the two hazards can cause larger damages than those caused by the either of hazards in isolation, and recent events have occurred in, e.g., Mozambique (2019), Texas (US, 2017), the Shoalhaven Estuary (Australia, 2016), Ravenna (Italy, 2015), Cork (Ireland, 2009), and Lymington (United Kingdom, 1999) (Couasnon et al., 2020; Zscheischler et al., 2018; Kumbier et al., 2018; Bevacqua et al., 2017; Olbert et al., 2017; Hendry et al., 2019). Practitioners and the scientific community are becoming more aware of CF risk, and there were several recent studies addressing the phenomenon at local (Bevacqua et al., 2017; Kumbier et al., 2018; van den Hurk et al., 2015) or larger scales ~~(Wahl et al., 2015; Bevacqua et al., 2019; Couasnon et al., 2020; Ward et al., 2018; Ganguli and Merz, 2019a, b)~~(Wahl et al., 2015; Bevac
. Recent advances in large-scale sea level and river discharge modelling allowed the generation of sub-daily time-series of water levels along the global coastline (Vousdoukas et al., 2018; Muis et al., 2016), thus enabling continental CF assessments (Ward et al., 2018; Bevacqua et al., 2019; Couasnon et al., 2020).

CF can be the result of different mechanisms depending on the local topography and meteorology. According to Wahl et al. (2015), CF is possible to occur when: (1) the joint occurrence of high river discharge and storm surge in estuarine regions may elevate water levels to a point where flooding is initiated or its impacts exacerbated; (2) a destructive storm surge, which already caused widespread flooding, is followed by rainfall, as the latter can drive additional flooding, even if it is not an extreme event on its own; and (3) a moderate storm surge occurs which does not directly cause flooding, but is high enough to fully block or slow down gravity-fed storm water drainage, and as a result precipitation causes flooding. In addition, CF may occur if (4) precipitation falls on wet soil that is saturated by a preceding storm surge (Bevacqua et al., 2019).

Quantifying the actual CF, i.e. the water level resulting from the combination of pluvial/fluvial flooding and high sea level, is challenging. For example, quantifying and interpreting CF in the vicinity of rivers (mechanism 1) requires water level measurements in the river mouth, which are rare, probably because most gauges are installed to monitor either riverine or marine processes (Bevacqua et al., 2017; Paprotny et al., 2018). Model-based data are also limited because only recently have modellers started considering CF. Statistical and hydrodynamic modelling approaches integrating fluvial and sea-level flooding have been developed and applied recently at the local scale ~~(Bevacqua et al., 2017; van den Hurk et al., 2015; Kumbier et al., 2018)~~ (Bevacqua et al., 2017; van den Hurk et al., 2015; Kumbier et al., 2018; Khanal et al., 2019); however, at large scale, these approaches are only now being developed. Similarly, explicit study of the actual CF water level due to pluvial flooding and storm surges (mechanisms 2-4) has not received much attention so far, to our knowledge, which may also be due to the scarcity of data. Thus, to gain information on the CF hazard at the regional, continental, or global scales, scientists usually employ proxies of flood hazard, e.g. the probability of potential CF (Bevacqua et al., 2019; Ward et al., 2018; Wahl et al., 2015; Couasnon et al., 2020; Paprotny et al., 2018). Quantifying potential CF, under the present or future climate, is useful as it allows identifying potential hotspots of CF hazard. Then, more detailed assessments of the local CF risk can be carried out at such hotspots, using computationally intensive methodologies that integrate all the hydrological and meteorological sources of flooding and their physical interaction (Wahl et al., 2015).

The large-scale assessment of potential CF includes the analysis of the probability (or return periods) of concurring extreme values of CF drivers. Two main approaches exist, focussing - respectively - on the analysis of the probability of concurring high values of sea level and river discharge, or of sea level and precipitation. Through considering river discharge, the first approach takes into account CF in estuaries and deltas, i.e. serving as a proxy of CF mechanism 1. The second approach, based on the analysis of accumulated precipitation around the coast when the high sea levels occur, can represent CF due to local precipitation extremes, i.e. related to mechanisms 2, 3, and 4. Given that precipitation is among the main drivers of river discharge, the two proxies are correlated to a certain extent and the use of precipitation can thereby allow quantifying CF potential also in certain river estuaries. However, the correlation between the two proxies can be sometimes poor, especially in locations where river discharge is strongly influenced by other factors such as snowmelt, evaporation, and accumulated precipitation over previous weeks or months (Blöschl et al., 2019).

Given the scarcity and heterogeneous distribution of in situ data (Ward et al., 2018; Couasnon et al., 2020; Wu et al., 2018), scientists have started to employ model data - of river, storm surge, and precipitation - to assess the large-scale potential CF hazard (Ward et al., 2018; Bevacqua et al., 2019; Wu et al., 2018; Couasnon et al., 2020; Wu et al., 2018; Paprotny et al., 2018; Bevacqua et al., 2020). Against the foregoing background, the present study aims to assess whether a precipitation based large-scale CF assessment can be used as a surrogate for potential CF in estuaries at the large-scale. To that end we use coherent global model datasets of storms surges (including wave effects) (Vousdoukas et al., 2018), precipitation (Beck et al., 2017b), and river discharge (Couasnon et al., 2020; Eilander, 2019) and conduct a first global comparison of the results obtained through the two approaches, keeping all the other methodological aspects identical.

2 Data

We analyse the period 1979-2015. We consider river discharge daily maxima from a publicly available global dataset (Eilander, 2019; Couasnon et al., 2020), which includes coastal catchments larger than 1,000 km^2. The dataset was the result of hydrological model simulations forced with temperature and potential evaporation derived from ERA-Interim, and with precipitation from the MSWEPv1.2 dataset (Couasnon et al., 2020). The latter is obtained by merging gauge, satellite, and reanalysis data (including ERA-Interim); more information can be found in Beck at al. (Beck et al., 2017b).

Precipitation is taken from the same MSWEPv1.2 dataset used to simulated river discharges and consists of daily data over a 0.25° grid. On each day we consider accumulated precipitation amounts within a 3-day centered window. This enables us to account for precipitation occurring just before and after midnight of the storm surge day (Bevacqua et al., 2019; Martius et al., 2016), and to consider different mechanisms causing CF (Wahl et al., 2015; Bevacqua et al., 2019).

Storm surges and waves were simulated with the hydrodynamic model DFLOW FM (Vousdoukas et al., 2017, 2018) and the wave model Wavewatch III (Mentaschi et al., 2017; Vousdoukas et al., 2017, 2018). The wave model was forced by 6-hourly wind, while DFLOW-FM was also forced by sea level pressure fields, both available from the ERA-Interim reanalysis (Dee et al., 2011). The effects of tropical cyclones (TCs) were considered in the reanalysis through storm surge simulations forced by downscaled atmospheric fields from all recorded TCs and by considering satellite-observed TC wave extremes (Vousdoukas et al., 2018). Astronomical tides are not considered in this analysis in order to focus on the meteorological component of the sea level, and excluding the stochastic coupling with tide induced water level variations. We consider daily water level maxima from the combined result of storm surges and wave setup (hereinafter mentioned as storm surges) according to Vousdoukas et al. (2018).

We analyse CF only around river mouth locations whose nearest precipitation and storm surge grid points lie within a distance of 75 km (Couasnon et al., 2020). This results in considering locations at river mouths of catchments with size in between about 1,000 and 3,690,000 Km^2 (95% having size smaller than 50,000 Km^2; Fig. 3f).

2.1 Methods

We assess the potential CF hazard via bivariate return periods of concurring extreme events ~~(Vandenberghe et al., 2011)~~ (Vandenberghe et al., 2011; Manning et al., 2019) of the variables X and Y, i.e. storm surge and precipitation ($\cancel{CF_{prec}}$$\mathrm{CF}_{\mathrm{prec}}$) or storm surge and river discharge ($\cancel{CF_{river}}$$\mathrm{CF}_{\mathrm{river}}$). Extremes of the individual variables ($\cancel{x_{ext}}$ ~~and~~ $\cancel{y_{ext}}$$x_{\mathrm{ext}}$ and y_{ext}) are defined as the associated α-year return levels. We use different α in the following, though we present the main results for α=5; images for α=2 are shown in the supplementary material. Return levels are obtained through fitting a generalized extreme value distribution to the annual maxima of the individual variables (Coles et al., 2001). Annual maxima are defined based on adjacent windows centred on the month where the climatological river discharge average is the highest, rather than from January-

December windows (such a window definition reduces the chance of selecting two consecutive annual maxima belonging to the same river discharge extreme event, and therefore leads to a more robust definition of the return levels). Overall, given the definition of the extremes based on α-year return levels, the bivariate return period is inherently linked to the dependence of the pairs in the tail of the distribution.

CF bivariate return periods are computed following the methodology presented by Bevacqua et al. (2019). The CF return period computation is based on the bivariate distribution of the variables of interest (X,Y), which is estimated semi-empirically to allow for robust estimation. For a given location, we select pairs whose individual values are simultaneously larger than the individual 95th percentiles (~~x_{sel} and y_{sel}~~x_{sel} and y_{sel}), and we model these pairs via a copula-based distribution. If the defined thresholds result in a small group of selected pairs, we lower the 95th percentile selection threshold to guarantee having ~~a~~ at least 20 pairs. The choice of 20 pairs is a trade off between having a sufficient amount of selected pairs and employing a reasonably high threshold for the fit of the parametric distribution in the tail. Furthermore, the return periods are largely insensitive to changes in the threshold (results are similar based on 20, ~~i. e. at least 20. Clusters of pairs~~ 30, and 40 pairs; not shown). The selection thresholds are generally high: 75% of the locations have a selected-threshold larger or equal to 0.95 and 0.94 for the precipitation- and river-based analysis, respectively. And 95% (99%) of the locations have a selected-threshold above 0.93 (0.885) and 0.89 (0.85) for the precipitation- and river-based analysis, respectively.

Once pairs are selected, clusters of pairs separated by less than 3 days were considered as part of the same event represented by the maximum X and Y values observed in the cluster. Note that while this choice has the drawback of not fully respecting the assumptions of independent realisations of the extreme events, which is necessary to apply extreme values theory in its generic form, it allows considering multiple storm surges that may occur during a sustained period of high river discharge and that could lead to multiple compound floods.

The return period is defined as:

$$\begin{aligned} T(x_{\mathrm{ext}}, y_{\mathrm{ext}}) &= \frac{\mu}{P((x > x_{\mathrm{ext}}\ and\ y > y_{\mathrm{ext}})\mid(x > x_{\mathrm{sel}}\ and\ y > y_{\mathrm{sel}}))} = \\ &= \frac{\mu}{1 - u_{\mathrm{Xext}} - u_{\mathrm{Yext}} + C_{\mathrm{XY}}(u_{\mathrm{Xext}}, u_{\mathrm{Yext}})} \end{aligned} \tag{1}$$

where μ is the average time elapsing between the selected pairs, ~~$u_{Xext} = F_X(x_{ext})$, F_X~~ $u_{\mathrm{Xext}} = F_{\mathrm{X}}(x_{\mathrm{ext}})$, F_{X} is the marginal cumulative distribution of the excesses over the selection threshold (accordingly for Y), and ~~C_{XY}~~ C_{XY} is the copula modelling the dependence between the selected pairs ~~.~~ (see Fig. A1 for the dependence associated with the fitted copulas in the two assessments). Note that, as the return period is obtained as a combination of the average elapsing time μ and the parametric probability density function of the data in the tail, an exact correspondence between the dependence of the copula and the return period is not expected. We model the marginal distributions of X and Y beyond the selection thresholds by a Generalised Pareto Distribution (GPD). We fit copulas from the families Gaussian, t, Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, BB8 to (u_X, u_Y) (obtained via empirical marginal cumulative distribution function ~~(Vandenberghe et al., 2011)~~ (Vandenberghe et al., 2011; Manning et al., 2018)); then we select the best ranked family according to the Akaike information criterion. In general, the physical processes captured by the two assessments can differ (even at the same location), therefore we allow for the selection of different copulas in the two assessments. We fit copulas and marginal distributions via a maximum likelihood estimator (using the *VineCopula* (Schepsmeier et al., 2016) and *ismev* (Heffernan et al., 2016) R-packages). We test the goodness of fit of copulas and marginals via the Cramer-von-Mises criterion (via the *eva* (Bader and Yan, 2016) and *VineCopula* (Schepsmeier et al., 2016) R-packages respectively).

When referring to the assessment of whether the CF return periods based on river discharge ($\cancel{T_{river}}$$T_{\mathrm{river}}$) are statistically different from those based on precipitation ($\cancel{T_{prec}}$$T_{\mathrm{prec}}$) or not, we use the concept of statistical compatibility, recently introduced by Amrhein et al. (Amrhein et al., 2019). We compute the ~~centred~~ centered 95% (2.5-97.5%) confidence interval of $\cancel{T_{prec}}$ T_{prec} on the basis of 600 resampled bivariate time series of precipitation and storm surge (each of them built randomly combining observed 1 calendar year bivariate time series (Bevacqua et al., 2020)). $\cancel{T_{river}}$ T_{river} is regarded as being statistically compatible with $\cancel{T_{prec}}$ ~~if~~ $\cancel{T_{river}}$ T_{prec} if T_{river} lies within the 95% confidence interval of $\cancel{T_{prec}}$$T_{\mathrm{prec}}$, and incompatible otherwise.

We qualitatively investigate how the two assessments compare for different classes of catchment size. To do so, we rank the rivers based on their catchment size and divide them into groups having the same sample size; for each group we compute different statistics to compare the two assessments: Spearman correlation of $\cancel{T_{river}}$ ~~and~~ $\cancel{T_{prec}}$~~, ratio~~ $\cancel{T_{river}/T_{prec}}$$T_{\mathrm{river}}$ and T_{prec}, ratio $T_{\mathrm{river}}/T_{\mathrm{prec}}$, and percentage of locations with $\cancel{T_{river}}$ ~~compatible to~~ $\cancel{T_{prec}}$$T_{\mathrm{river}}$ compatible to T_{prec}. This binning procedure provides equally robust statistics for each bin and shows similar results for small variations in the bin size.

2.1.1 Results and discussion

The spatial patterns of the potential CF return periods based on either precipitation ($\cancel{T_{prec}}$$T_{\mathrm{prec}}$), or river discharge ($\cancel{T_{river}}$$T_{\mathrm{river}}$) are very similar (Fig. 1; Fig. ~~A1~~ A2 is identical but shows results based on extremes defined considering 2-year return levels). The results for clusters of locations with the 5% lowest CF return periods are also similar in the two assessments (Fig. ~~A4 and A5~~A3 and A4). These hotspot regions are mainly found along the US and central American coasts, Central Chile, Madagascar, the southern North Atlantic coasts, and southern Japan.

While the spatial patterns of the CF return periods obtained from the two approaches are very similar, their relative differences can be substantial, especially at certain locations (Fig. 2 and ~~A2~~A5). Given that the return period computation procedure involves several uncertainty factors (e.g. bivariate model fitting and definition of the return levels), we test the hypothesis that the return period based on the river discharge is statistically compatible (at 95% confidence level) with that based on precipitation. When defining extremes based on the 5-year return levels, the river-based return period is compatible with the precipitation-based value in about 82% of the locations (Fig. 2c; 76% for 2-year return levels: Fig. ~~A2c).~~ A5c). The spatial distribution of locations where $\cancel{T_{river}}$ T_{river} is not statistically compatible with $\cancel{T_{prec}}$ T_{prec} does not seem to follow a clear spatial pattern, though it appears that $\cancel{T_{river}}$ T_{river} is lower than $\cancel{T_{prec}}$ T_{prec} in northern Europe and in the tropics. The latter are areas where CF is unlikely. $\cancel{T_{river}}$ T_{river} is higher than $\cancel{T_{prec}}$ T_{prec} along the Gulf of Mexico (Fig. 2c and Fig. ~~A2c~~A5c). Compatible $\cancel{T_{river}}$ ~~and~~ $\cancel{T_{prec}}$ T_{river} and T_{prec} are found but with large discrepancies in the tropics and above ~~60deg~~ 60^o North (Fig. 2b), consistently with the high uncertainty of these large CF return periods that do not allow for detecting potential differences between $\cancel{T_{river}}$ ~~and~~ $\cancel{T_{prec}}$~~.~~T_{river} and T_{prec}.

In areas with a tendency towards high CF return periods, e.g. the tropics, neighbour locations show divergent values in the ratio between the return periods of the two assessments (dark blue and red dot in Fig. 2). Further tests showed that this behaviour

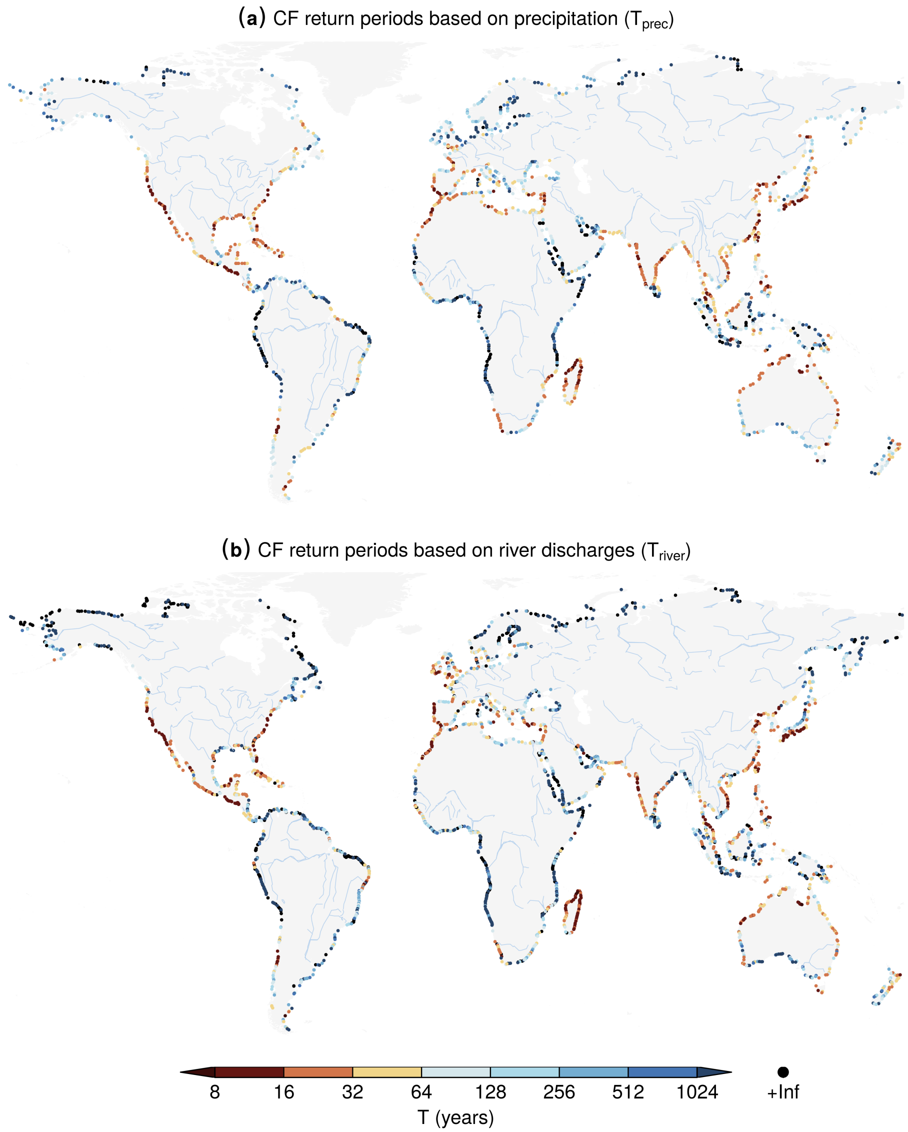


Figure 1. Present-day (1980-2014) potential compound flooding probability based on precipitation and on river discharge. Return periods of CF defined as co-occurring extremes (5-year return levels) of the CF drivers. Return period of co-occurring ~~(a)~~ **(a)** storm surge (including waves) and precipitation (accumulated 3-day centred) extremes; ~~(b)~~ **(b)** storm surge (including waves) and river discharge extremes. Major rivers are shown in light blue.

is not related to goodness of fit of the bivariate distributions, rather it appears associated with the large uncertainties of high return periods and potentially with different catchment characteristics.

We find that there is a tendency towards higher differences in the two assessments at locations where either or both T_{prec} and T_{river} are high (i.e. where T approaches the value expected under independence of the CF drivers) (Fig. A6). (This appears consistent with the high uncertainty associated with large CF return periods.) Such a finding has relevant implications, as it indicates that the two assessments tend to be similar, on average, where assessing the actual CF is more important, i.e. where there is a relatively high CF potential (Fig. A6).

The spatial association of the return periods map obtained from the two assessments is shown in Fig. 3a and ~~A3a~~A7a. The Spearman correlation between ~~T_{river} and T_{prec}~~ T_{river} and T_{prec} is $\sim$0.7, and increases as the return level threshold employed to define the extremes decreases; e.g., 0.66 and 0.73 for the 5- and 2-year return level, respectively (Fig. 3a and ~~A3a~~A7a). This trend in the correlation is consistent with the higher uncertainties characterizing larger return periods.

Although some of the differences between the two assessments ~~may~~ can be driven by uncertainties in the return period estimation, also several physical processes, e.g., topography-dependent, may be involved. The Spearman correlation between the two assessments decreases (p-value < 0.0024) as the catchment size increases (Fig. 3c). The latter suggests that - for larger catchments - different processes may cause either positive or negative deviations between ~~T_{river} and T_{prec}~~ T_{river} and T_{prec}.

Water levels in the mouth of small rivers are expected to be largely influenced by the precipitation around the coast (Hendry et al., 2019; van den Hurk et al., 2015; Bevacqua et al., 2017), while around large rivers inland hydrological processes are usually dominant. Therefore, we qualitatively investigate whether the actual difference between the two assessments (quantified as ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$) depends on the size of the catchment (Hendry et al., 2019). For example, the ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ ratio, defined using the 5-year return level, tends to increase with the catchment size (Fig. 3b; see the deviation of the scatter plot of ~~T_{river} vs T_{prec}~~ T_{river} vs T_{prec} from the identity line emerging with the increasing dimension of the catchment). For CF return periods defined based on 2-year return level extremes the median ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ is near unity for most of the catchments, even though ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ increases slightly for larger catchments (blue line in Fig. 3d). For higher return levels, the median ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ increases with the size of the catchment (Fig. 3d), indicating that the CF assessments based on precipitation and river discharge differ largely for large rivers. For all return levels, the median ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ is near unity for rivers whose catchment size is up to about 5-10,000 Km^2 (Fig. 3d), i.e. about 75% of the rivers presently analysed (Fig. 3f). Binning the rivers per catchment size we find that the variance of the ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ ratio within each bin tends to be smaller for small than for large rivers (Fig. 3d, shown by the dashed line for the 2-year return level). However, it is important to highlight that there is substantial variance (Fig. 3d for the 2-year return level extremes), and that for a river of any catchment size, the associated ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ can be either small or large. Overall, the similarity of ~~T_{river} and T_{prec}~~ T_{river} and T_{prec} for small rivers is highlighted by the fact that for small rivers it is more likely that ~~T_{river}~~ T_{river} is statistically compatible with ~~T_{prec}~~ T_{prec} (Fig. 3e). For example, Fig. 3e (4-year return level curve) shows that ~~T_{river}~~ T_{river} is statistically compatible with ~~T_{prec}~~ T_{prec} for $\sim$82% of the small rivers (catchment size $<$ 5,000 Km^2) and for $\sim$75% of large rivers (catchment size $>$ 50,000 Km^2). The decrease in the compatibility of ~~T_{river} and T_{prec}~~ T_{river} and T_{prec} with the catchment size (Fig. 3e) is statistically significant for return levels smaller than the 6-year return level (p-value < 0.022). However, for high return levels this decrease is not prominent, most likely due to the large uncertainty associated with longer return periods. Fig. 3e indicates the discrepancy for large catchments in ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ being greater for low return levels, whereas Fig. 3d indicates the opposite; this is also likely caused by the larger uncertainty of return periods of

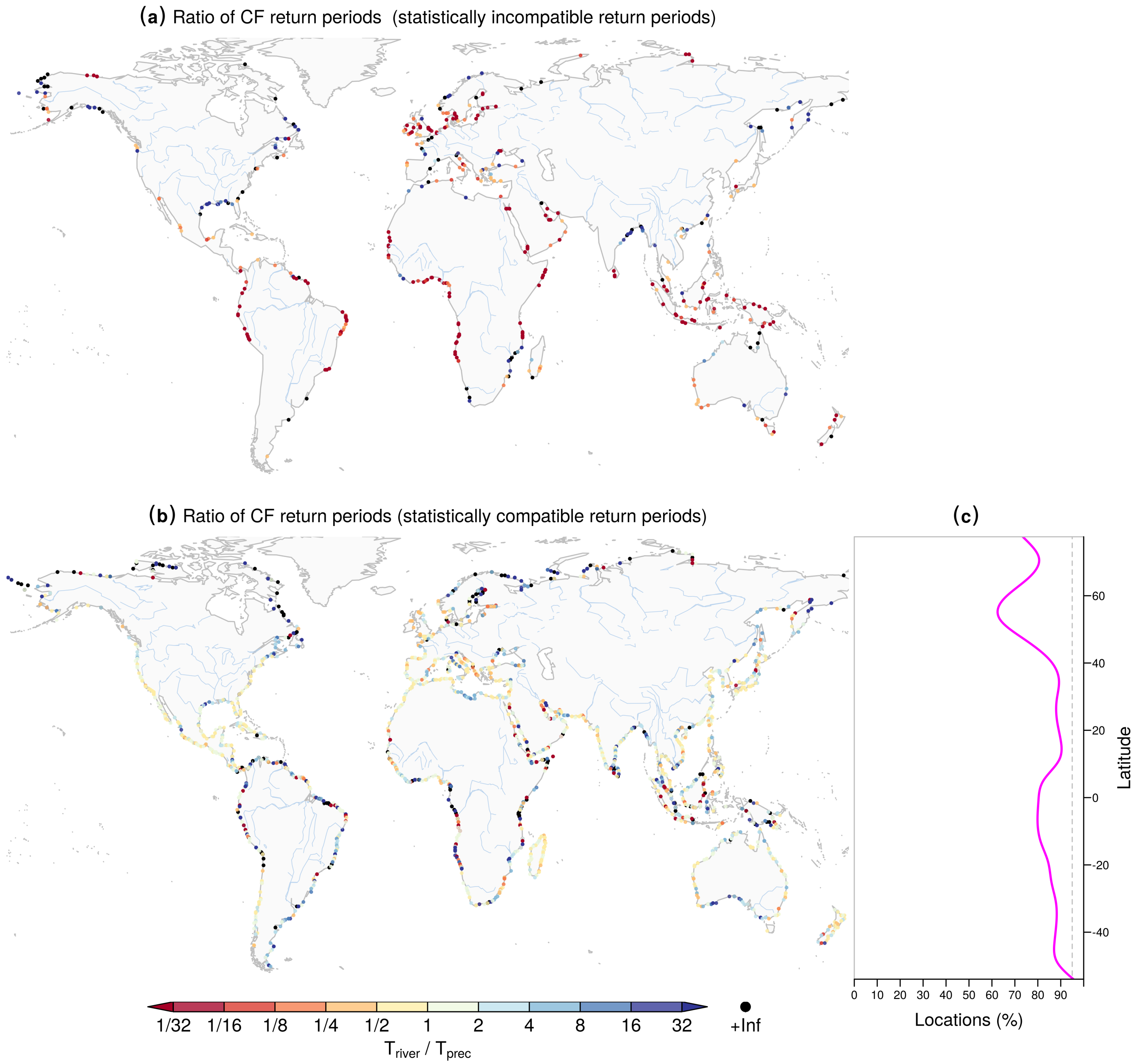


Figure 2. Difference between the present-day (1980-2014) potential compound flooding probability based on precipitation and on river discharge. Ratio (~~T_{river}/T_{prec}~~$T_{\mathrm{river}}/T_{\mathrm{prec}}$) of CF return period of concurring river discharge and storm surge extremes (~~T_{river}~~T_{river}), to CF return period of concurring precipitation (accumulated 3-day centred) and storm surge extremes (~~T_{prec}~~T_{prec}). Panel ~~(a)~~ **(a)** shows the ratio where ~~T_{river}~~T_{river} is statistically incompatible (at 95% confidence level, see methods) with ~~T_{prec}~~T_{prec}. Panel ~~(b)~~ **(b)** shows the ratio where ~~T_{river}~~T_{river} is statistically compatible with ~~T_{prec}~~T_{prec}, while panel ~~(c)~~ **(c)** shows the coastline fraction where this happens (binned every 5° of latitude and smoothed using a spline function). In (c), the dashed grey line shows the 95% level.

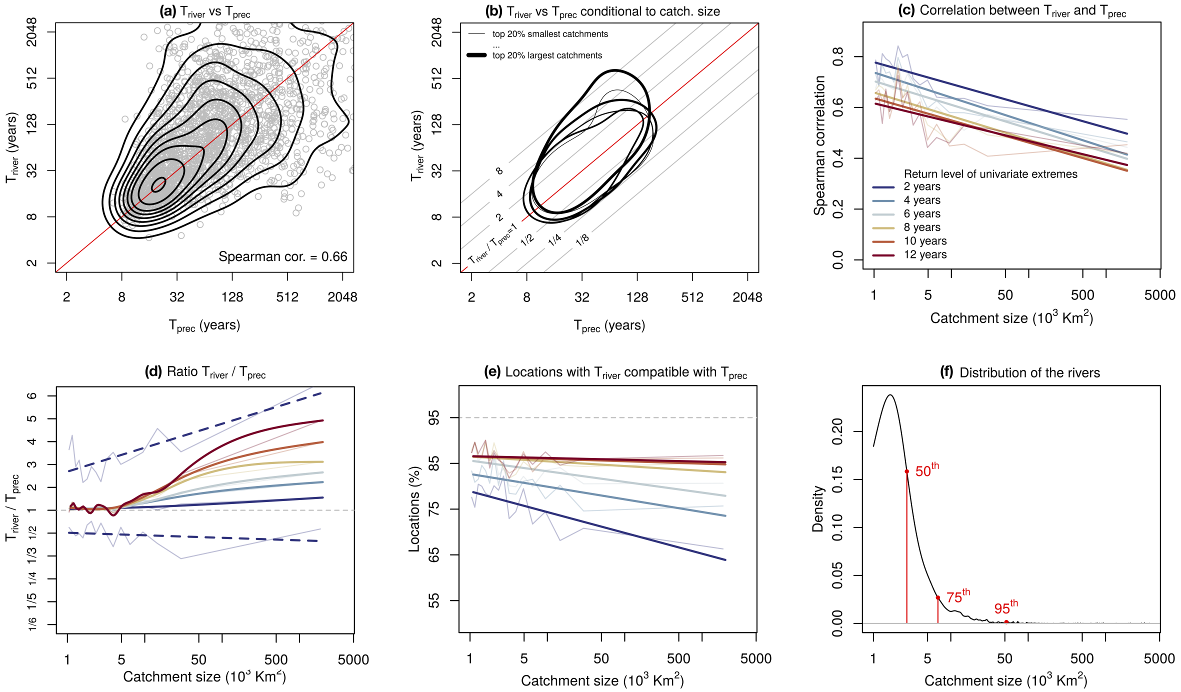


Figure 3. Comparison of the potential compound flooding probability based on precipitation and on river discharge. ~~(a)~~ **(a)** Scatterplot (based on Fig. ~~A1~~ A2) of CF return periods based on river discharge (~~T_{river}~~ T_{river}) and on precipitation (~~T_{prec}~~ T_{prec}), for extremes defined based on 5-year return levels. Black contours represent the isolines of the kernel density containing from 10% to 90% of the (~~T_{river}, T_{prec}~~ $T_{\mathrm{river}}, T_{\mathrm{prec}}$) pairs. The identity line is shown in red. ~~(b)~~ **(b)** Empirical probability density function of the variables (~~T_{river}, T_{prec}~~ $T_{\mathrm{river}}, T_{\mathrm{prec}}$) conditioned on the catchment size; the thickness of the isolines increases with the catchment size, such that each bin considers 1/5 of the total number of analysed rivers (3178). Each contour line (isolines of the kernel density) contains about 50% of the (~~T_{river}, T_{prec}~~ $T_{\mathrm{river}}, T_{\mathrm{prec}}$) pairs. ~~(c-e)~~ **(c-e)** Statistics comparing ~~T_{river}~~ T_{river} and ~~T_{prec}~~ T_{prec} as a function of the catchment size and of the return levels used to define the CF univariate extremes (see legend in panel (a)). ~~(c)~~ **(c)** Spearman correlation between the maps of ~~T_{river}~~ T_{river} and ~~T_{prec}~~ T_{prec}. ~~(d) T_{river}/T_{prec}~~ **(d)** $T_{\mathrm{river}}/T_{\mathrm{prec}}$ ratio. The dashed line is the centered 68% (16th-84th percentiles) confidence interval of the ratio based on the 2-year return levels. A non linear Y-axis for values below 1 is plotted such that the specular cases, e.g., ratio r=2 (~~$T_{river} = 2 \cdot T_{prec}$~~ $T_{\mathrm{river}} = 2 \cdot T_{\mathrm{prec}}$) and r=1/2 (~~$T_{prec} = 2 \cdot T_{river}$~~ $T_{\mathrm{prec}} = 2 \cdot T_{\mathrm{river}}$) (see Fig. 3b) appear symmetric with respect to r=1 (~~$T_{river} = T_{prec}$~~ $T_{\mathrm{river}} = T_{\mathrm{prec}}$). (To obtain the plot, the ratio is defined as ~~$r = T_{river}/T_{prec}$~~ $r = T_{\mathrm{river}}/T_{\mathrm{prec}}$; then if r<1, r is transformed to ~~$r = -T_{prec}/T_{river} + 2$~~ $r = -T_{\mathrm{prec}}/T_{\mathrm{river}} + 2$. Now, e.g., r=2 and r=0 represent specular cases, therefore, where the y-axis is r=0,1,2,..., we can write r=1/2,1/3,1/4,...). See Fig. ~~A3b~~ A7b for the plot with standard axis. ~~(e)~~ **(e)** Percentage of locations where ~~T_{river}~~ T_{river} is statistically compatible with ~~T_{prec}~~ T_{prec} at the 95% confidence level. ~~(f)~~ **(f)** Probability density function of the catchment size of the analysed rivers; 50th, 75th, and 95th percentiles of the distribution are shown in red. In ~~(c-e)~~ **(c-e)**, tick lines are obtained through regressing the ~~investigated~~ obtained statistical values per bin (thin lines) to the natural logarithm of the mean size of the catchment bins. All lines are regressed using linear regression, except from the median of the ratio in ~~(d)~~ **(d)** where a spline function is used. The slopes of the linear regressions are all significant in **(c)** (~~c)~~ (p-value < 0.0024), and significant up to the 6-year return level in **(e)** (~~e)~~ (p-value < 0.022).

higher return levels, that does not allow for detecting potential differences between large values of ~~T_{river} and T_{prec}~~T_{river} and T_{prec}.

The differences between ~~T_{river} and T_{prec}~~ T_{river} and T_{prec} are not only controlled by the catchment size, but can be the result of several other factors. During the hydrological modelling, input data artefacts, or model inaccuracies, among others, introduce uncertainty which may contribute to the observed differences. Another important contribution should arise from the uncertainties in the bivariate return period estimation (Bevacqua et al., 2019; Wahl et al., 2015), which can contribute to both positive and negative deviations between ~~T_{river} and T_{prec}~~T_{river} and T_{prec}. Moreover, the catchment response time to precipitation depends also on rock and soils catchment permeability (Hendry et al., 2019). Finally, river discharge is influenced not only by local coastal precipitation, but also by the weather over the previous weeks or months over the catchment including evaporation and potentially snowmelt (Couasnon et al., 2020; Bevacqua et al., 2017).

Clearly, the diversity of the physical processes leading to ~~CF_{prec} and CF_{river}~~ $\mathrm{CF}_{\mathrm{prec}}$ and $\mathrm{CF}_{\mathrm{river}}$ is very relevant and can cause both positive and negative differences between ~~T_{river} and T_{prec}~~T_{river} and T_{prec} (Blöschl et al., 2019). For example, for any given catchment, a slow catchment response time may either increase or decrease the ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ ratio. In locations where cyclones cause frequent concurring storm surge and widespread coastal precipitation, it is not guaranteed that ~~CF_{river}~~ $\mathrm{CF}_{\mathrm{river}}$ probability will be also high. For example, if the rainfall in the catchment upstream needs time to reach the coast, long enough for the storm surge to recede, then ~~CF_{river}~~ $\mathrm{CF}_{\mathrm{river}}$ will be unlikely and in this case the ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ ratio will be high (Klerk et al., 2015; Ward et al., 2018). In contrast, where the ~~CF_{prec}~~ $\mathrm{CF}_{\mathrm{prec}}$ is unlikely because different weather systems cause precipitation and storm surge extremes, a relatively slow catchment response time may sometimes allow for high river discharge and storm surge to concur if, e.g., a first cyclone causes precipitation driving high river discharge reaching the coast when a second cyclone drives a storm surge (contributing to low ~~T_{river}/T_{prec}~~$T_{\mathrm{river}}/T_{\mathrm{prec}}$) (Bevacqua et al., 2019).

In addition the presence of a pronounced annual cycle in river discharge can modulate the river discharge driven CF hazard, and thus the ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ ratio. In regions where ~~CF_{prec}~~ $\mathrm{CF}_{\mathrm{prec}}$ is likely, ~~CF_{river}~~ $\mathrm{CF}_{\mathrm{river}}$ may be unlikely if the storm surge season does not coincide with the season of the high river discharge (Ward et al., 2018) (high ~~T_{river}/T_{prec}~~$T_{\mathrm{river}}/T_{\mathrm{prec}}$). In contrast, where the ~~CF_{prec}~~ $\mathrm{CF}_{\mathrm{prec}}$ is unlikely, ~~CF_{river}~~ $\mathrm{CF}_{\mathrm{river}}$ may be more likely if the storm surge season coincides with the high river discharge season (low ~~T_{river}/T_{prec}~~$T_{\mathrm{river}}/T_{\mathrm{prec}}$). In addition, precipitation extremes are typically short in duration, while river discharge extremes are less dynamic. Although events exceeding the α-year return level threshold occur both for precipitation and river discharge on average every α years, the number of days with high river discharge can be larger than the number of days with high precipitation amount. From a statistical point of view, the above mechanism alone would make ~~CF_{river}~~ $\mathrm{CF}_{\mathrm{river}}$ more likely than ~~CF_{prec} (low T_{river}/T_{prec}~~$\mathrm{CF}_{\mathrm{prec}}$ (low $T_{\mathrm{river}}/T_{\mathrm{prec}}$). (This effect is even more pronounced in catchments with a slow response time and especially in areas where different weather systems cause precipitation and storm surge extremes (Bevacqua et al., 2020).) The above effect is weakened as the return level threshold used to define extremes increases; due to the shorter duration of river discharge extremes, the above considerations justify ~~T_{river}/T_{prec}~~ $T_{\mathrm{river}}/T_{\mathrm{prec}}$ increasing with the return level choice (Fig. 3d).

While the relevance of these mechanisms may depend on the local climate, they are expected to be less relevant in very small rivers where the catchment response time is small, and thus autocorrelation in the river time series is smaller. Consistently, we find that ~~T_{river} and T_{prec}~~ T_{river} and T_{prec} tend to match more in smaller catchments. However, a higher agreement for small rivers might also arise from the relatively coarse spatial resolution of the data which would attenuate differences between precipitation and river discharges in small rivers. Overall, we find that ~~independently~~ independent of the catchment size, ~~T_{river}~~ T_{river} tends to be higher than ~~T_{prec}~~ T_{prec} on average (Fig. 3d), suggesting that the mechanisms causing ~~$T_{river} > T_{prec}$~~ $T_{\mathrm{river}} > T_{\mathrm{prec}}$ may be more likely or relevant. Apparently this is a general remark and not a universal law since there are also several locations where ~~T_{prec} exceeds T_{river}~~ T_{prec} exceeds T_{river}.

The presented results are based on state-of-the-art model data which have been validated and discussed in previous papers (Vousdoukas et al., 2018; Beck et al., 2017a; Couasnon et al., 2020; Bevacqua et al., 2019; Muis et al., 2016). However, the above mentioned validation efforts did not include areas where field measurements are scarce, like parts of South America, Africa, and Asia (Ward et al., 2018). At high latitudes, ice and snow dynamics apply a certain control to both river hydrology (Yamazaki et al., 2014) and wave and storm surge dynamics (Vousdoukas et al., 2017). However, such processes are not resolved by the global models used to generate the forcing datasets used in the present study. For these reasons, the present findings should be interpreted with care, especially in northern regions (Couasnon et al., 2020).

The two approaches investigated here provide information only on the *potential* for CF. The actual CF occurrence depends also on the local topography which can favour or not the interaction between the CF drivers; also, concurrent but not hydrologically-interacting storm surge and pluvial or fluvial flooding are relevant as they can, e.g., limit the ability to respond to emergency, and amplify the impacts that the two hazards would have caused if they occurred in isolation (Martius et al., 2016; Barton et al., 2016; Zscheischler et al., 2019). Moreover, while the two approaches investigated here are supposed to represent different CF mechanisms, separating the CF mechanisms in this way could be misleading, as CF may happen due to a combination of river discharge, local rainfall and associated surface runoff, together with high sea level. For example, in July 2019, New Orleans (Louisiana, US) (Vahedifard et al., 2016) was simultaneously threatened by the tropical Storm Barry causing local rainfall and storm surge around the coast, and by extremely high water discharge from the Mississippi River which lasted from March to July. Local CF risk assessment at the local scale should be carried out via complex hydrological modelling that can take into account the complex mechanisms causing CF, including storm surges, waves, astronomical tides, and when necessary not only fluvial or pluvial flooding but also their combination.

3 Conclusions

We conduct two global-scale potential CF hazard assessments, using either storm surge and precipitation, or storm surge and river discharge model data, and compare how the choice of precipitation versus river discharge as covariate with storm surge affects the results. We find that the two approaches result in similar spatial CF hazard patterns, which tend to deviate as the river catchment size increases. In addition, on average the deviations between the two assessments are smaller in regions where assessing the actual CF is more relevant, i.e. where the CF potential is high.

Due to data scarcity, current large-scale CF assessments rely on approaches and model-based datasets similar to those used here. This study indicates that for these large-scale assessments, a precipitation-based CF analysis can provide satisfactory information on the CF potential in estuaries of small and medium size rivers (catchment smaller than about 5-10,000 Km^2). Moreover, a precipitation-based CF analysis allows for assessing the CF hazard arising from the interaction of local coastal rainfall and storm surges where no rivers exist, or along the mouths of small rivers often not represented in global river datasets. Naturally, employing river discharge data should always be preferred to using precipitation when studying both the large and local scale CF in estuaries, when data are available, especially in areas where we detected large differences between the two approaches. The importance of using river discharge data is even greater in estuaries of long rivers.

Data availability. Precipitation data is available on request online (https://ec.europa.eu/jrc/en/publication/mswep-3-hourly-025-global-gridded-precipitation-1979-2014-merging-gauge-satellite-and-reanalysis). Sea level data are available at https://data.jrc.ec.europa.eu/collection/liscoast (further inquiries should be addressed to MIV). River discharge data were obtained from the dataset "Paired time series of daily discharge and storm surge" are available at: https://doi.org/10.5281/zenodo.3258007.

~~The same as Fig. 1, but the extremes are defined as 2-year return levels.~~

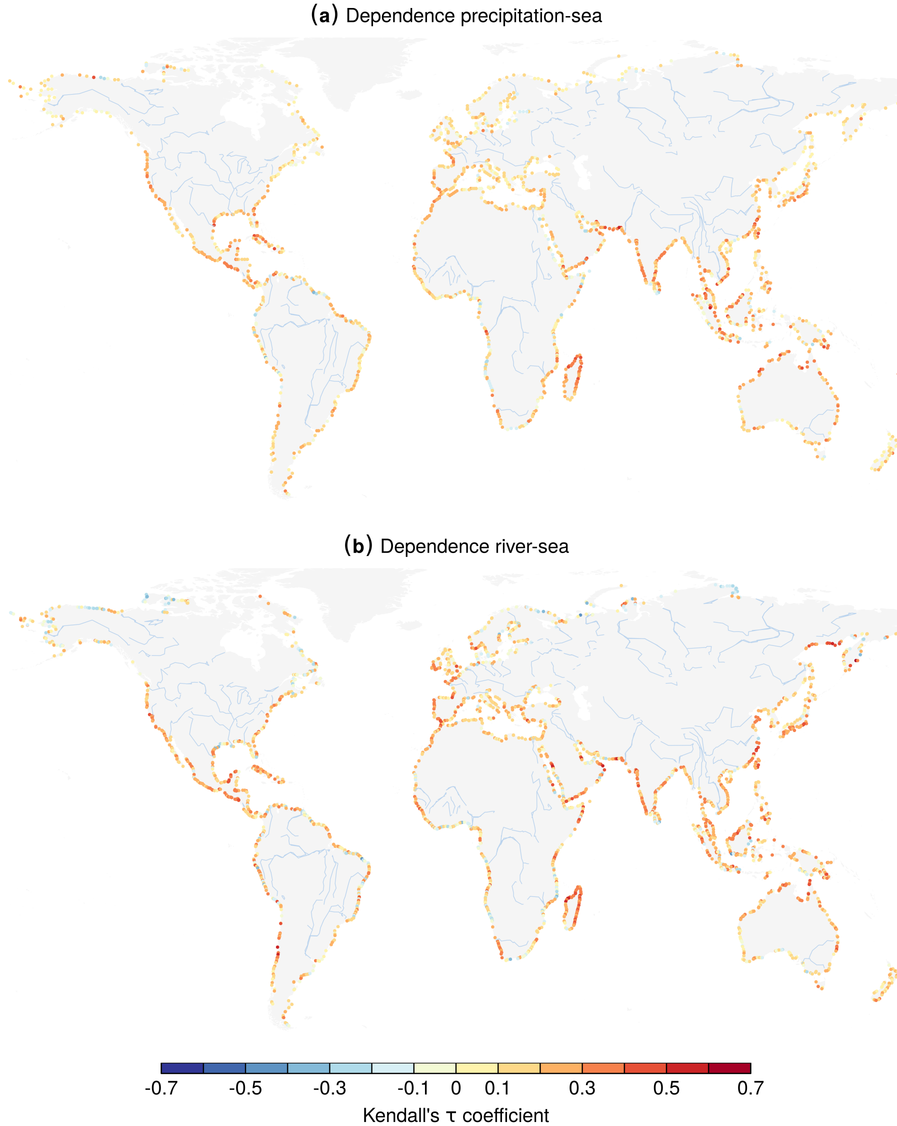
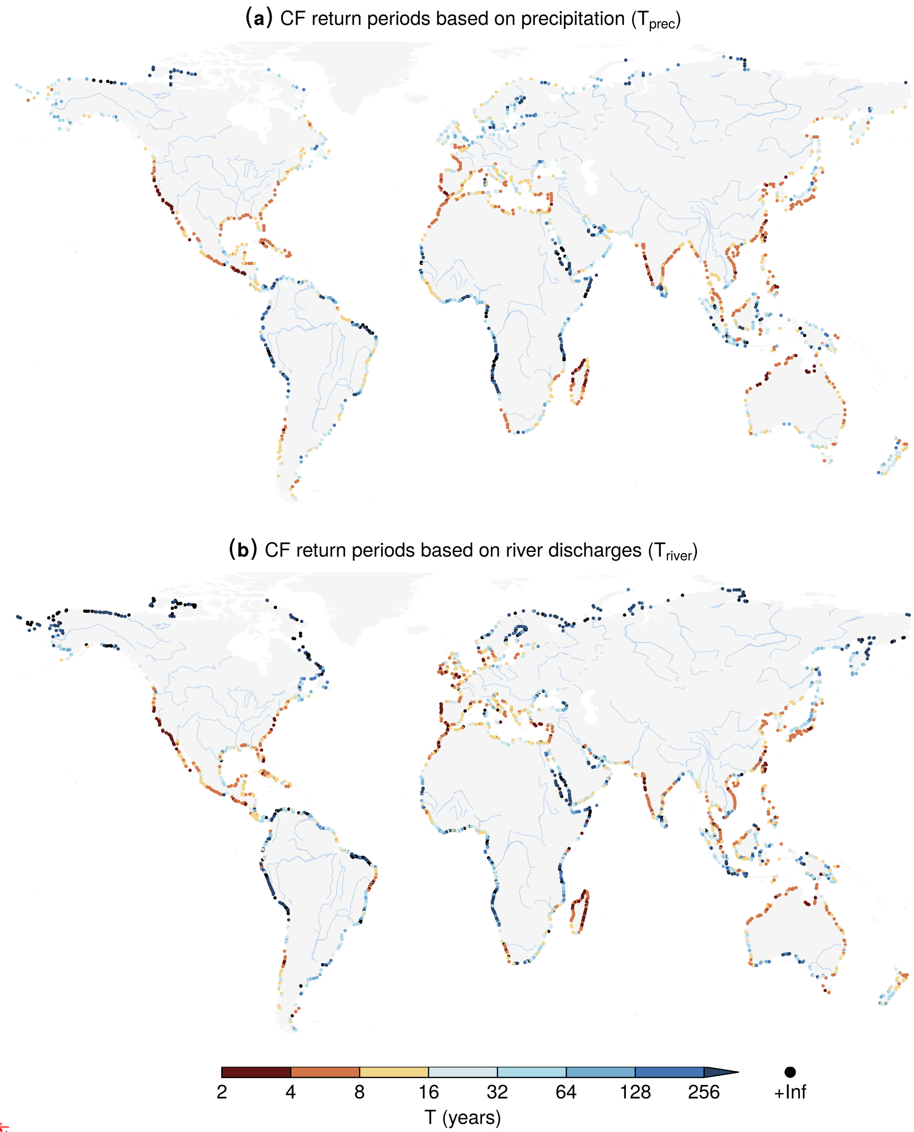


Figure A1. Kendall's tau correlation associated with the copulas fitted to the selected pairs of (a) precipitation and sea level, and (b) river discharge and sea level.

Appendix A: Supporting figures

~~The same as Fig. 1, but the results are based on extremes defined considering 2-year return~~



~~levels.~~

Figure A2. The same as Fig. 1, but the extremes are defined as 2-year return levels.

~~(a) The same as Fig. 3a, but the results are based on extremes defined considering 2-year return levels (the figure is based on Fig. A1). (b) As Fig. 3d, but with linear y-axis.~~

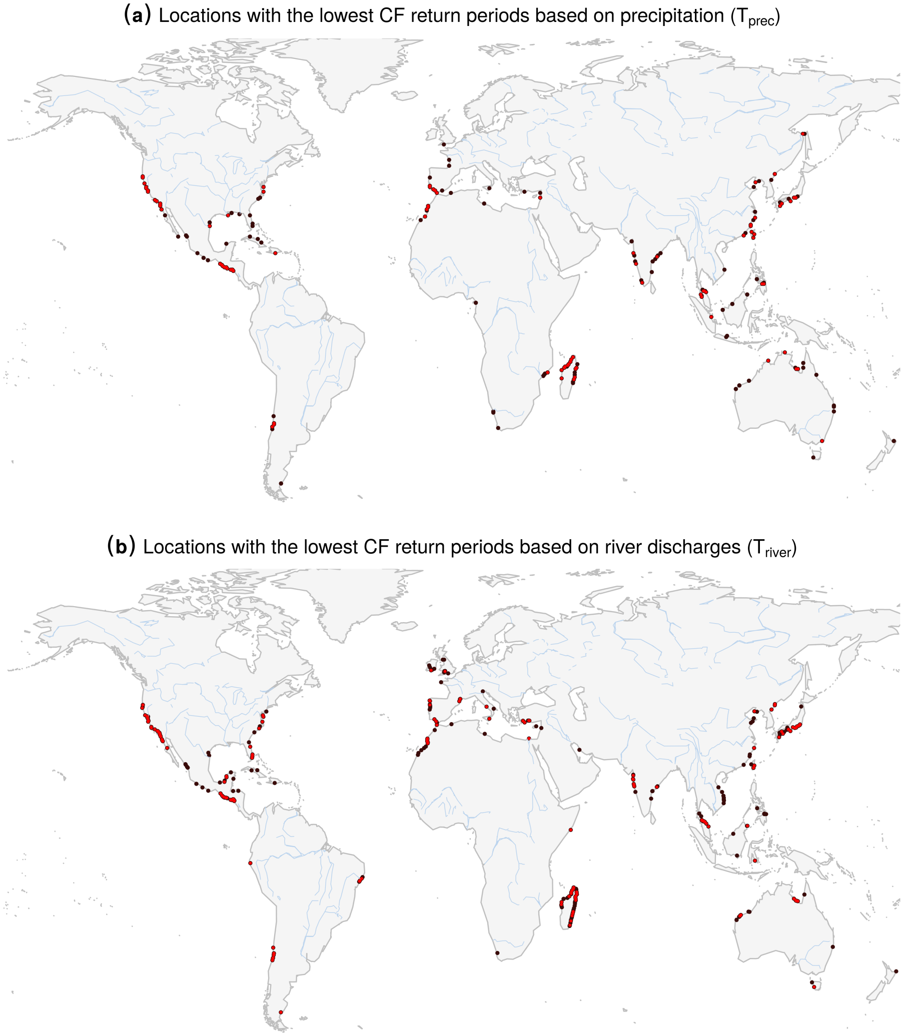


Figure A3. Locations with the lowest potential compound flooding probability based on precipitation and on river discharge. Locations with return periods below the 10th and 5th percentile are shown in black and red, respectively. The extremes are defined considering 5-year return levels. CF return periods are based on precipitation in ~~(a)~~ **(a)** and on river discharge in ~~(b)~~**(b)**.

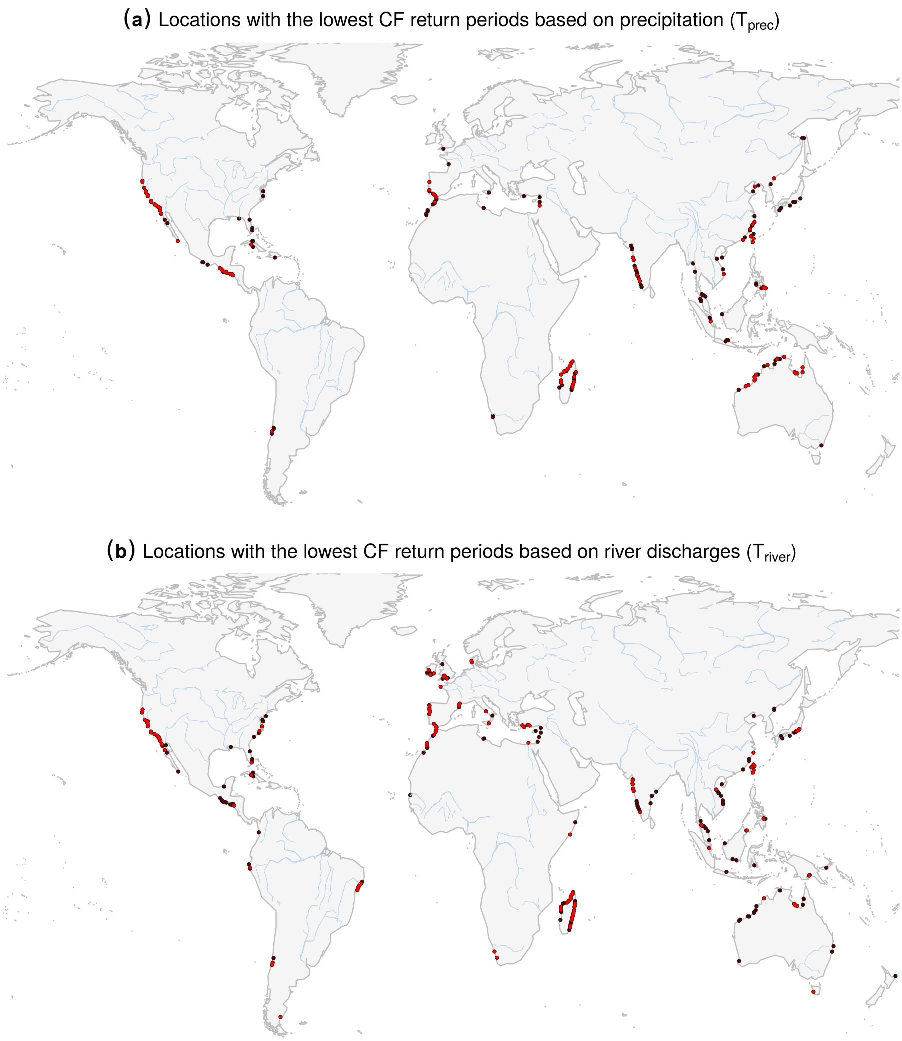


Figure A4. The same as Fig. ~~A4~~A3, but the results are based on extremes defined considering 2-year return levels.

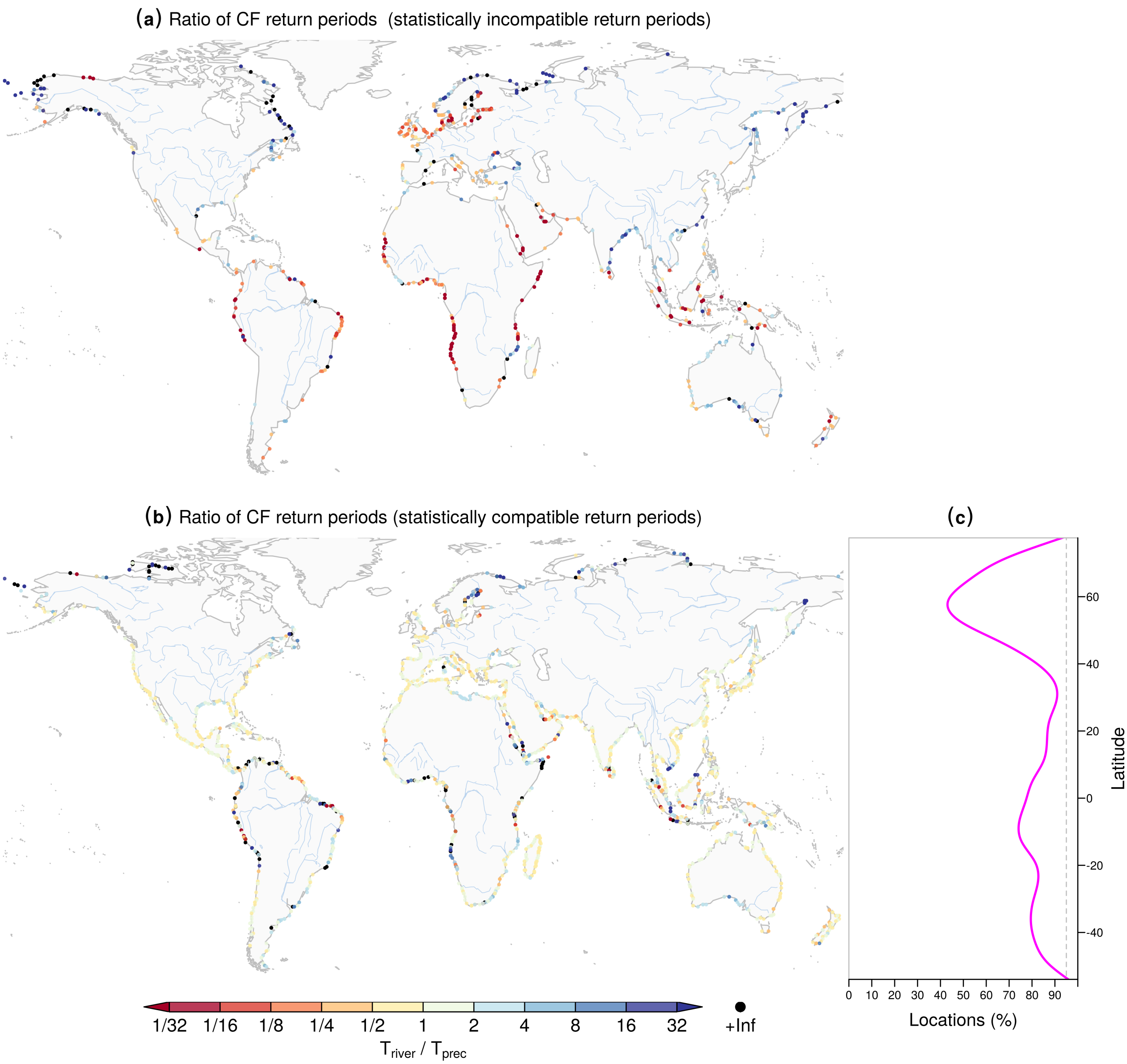


Figure A5. The same as Fig. 2, but the results are based on extremes defined considering 2-year return levels.

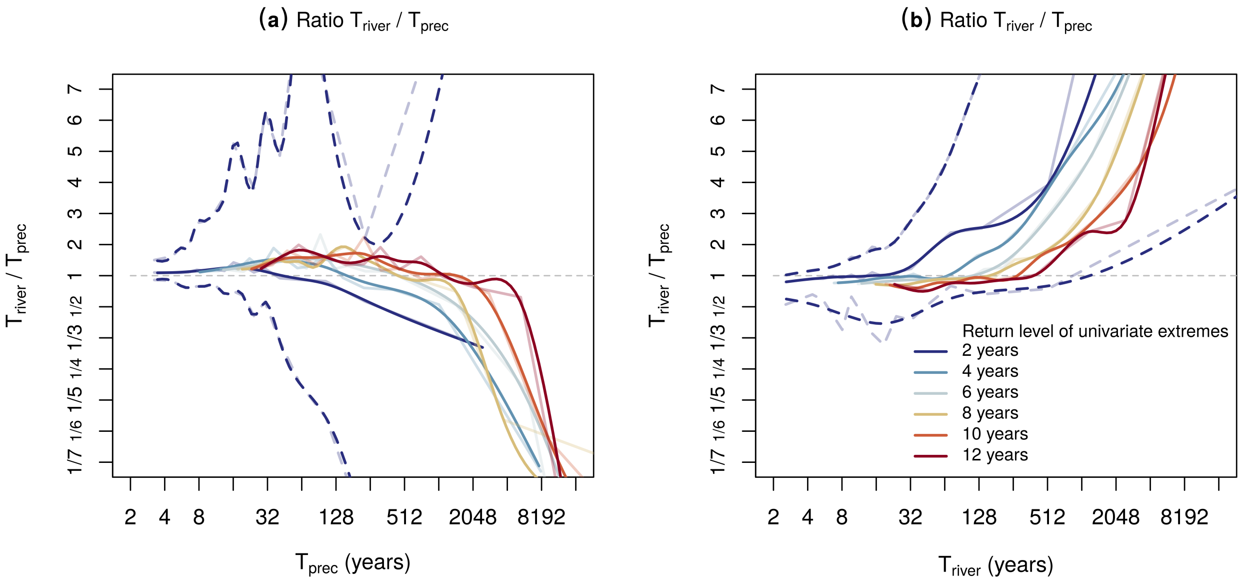


Figure A6. (a) Ratio between the CF return periods based on river discharge (T_{river}) and on precipitation (T_{prec}), as a function of T_{prec} (and of the return levels used to define the CF univariate extremes; see legend in panel (b)). The dashed line is the centered 68% (16th-84th percentiles) confidence interval of the ratio based on the 2-year return levels. Tick lines are obtained through regressing the investigated statistical values to the natural logarithm of the mean return period of the bins via a spline function. (b) The same as (a), but the ratio is conditioned on T_{river}. As in Fig. 3d of the original manuscript, a non linear Y-axis for values below 1 is employed such that the specular cases, e.g., ratio r=2 ($T_{river} = 2 \cdot T_{prec}$) and r=1/2 ($T_{prec} = 2 \cdot T_{river}$) (see Fig. 3b) appear symmetric with respect to r=1 ($T_{river} = T_{prec}$).

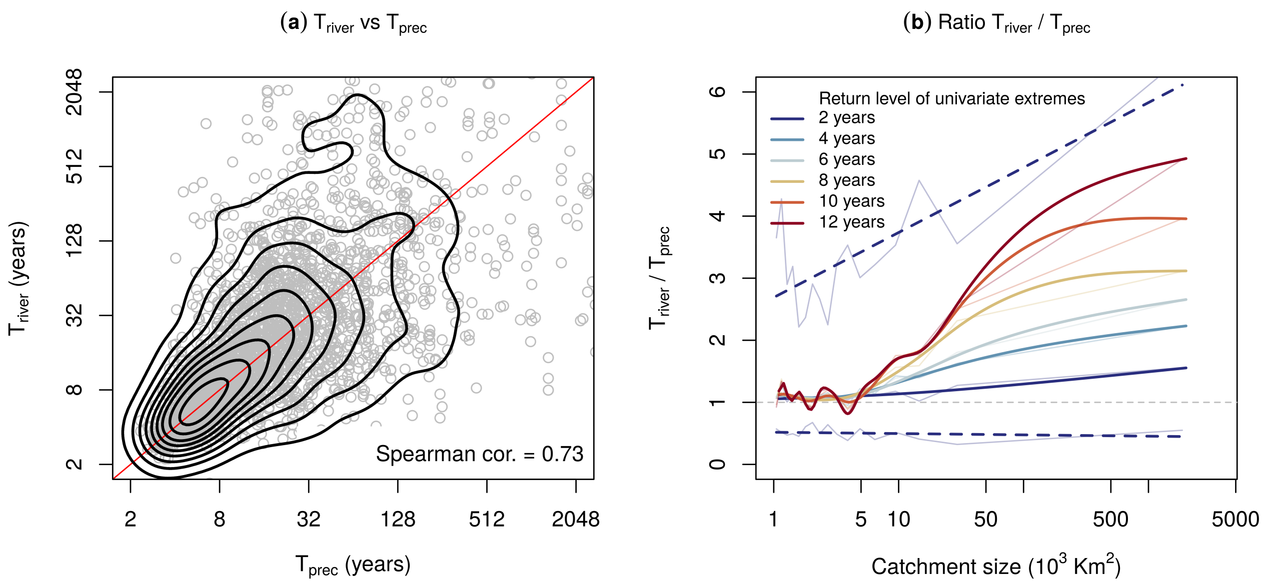


Figure A7. **(a)** The same as Fig. 3a, but the results are based on extremes defined considering 2-year return levels (the figure is based on Fig. A2). **(b)** As Fig. 3d, but with linear y-axis.

Author contributions. EB initiated the study, carried out the data analysis, and drafted the manuscript. EB designed the study development with contributions from MIV. EB and MIV worked on the final manuscript version. MIV performed the storm surge runs. All the authors gave conceptual inputs during the data analysis, discussed the results, and commented on the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. EB acknowledges financial support from the European Research Council grant ACRCC (project 339390). EB acknowledge the European COST Action DAMOCLES (CA17109). MV acknowledges financial support from the EUPHEME project and CoCliServ project, which are part of ERA4CS, an ERA-NET initiated by JPI Climate and co-funded by the European Union (grant no. 690462). The authors would like to thank Lorenzo Mentaschi and Evangelos Voukouvalas who contributed in the generation of the storm surge time series, and Dirk Eilander for creating and publicly sharing the dataset “Paired time series of daily discharge and storm surge” which was fundamental for developing this study. Finally, the authors would like to thank two anonymous reviewers for their valuable/constructive comments and suggestions that contributed to improving the manuscript.

References

Amrhein, V., Greenland, S., and McShane, B.: Scientists rise up against statistical significance, 2019.

Bader, B. and Yan, J.: eva: Extreme Value Analysis with Goodness-of-Fit Testing, R package version, 0.2.4, 2016.

Barton, Y., Giannakaki, P., Von Waldow, H., Chevalier, C., Pfahl, S., and Martius, O.: Clustering of regional-scale extreme precipitation events in southern Switzerland, Monthly Weather Review, 144, 347–369, 2016.

Beck, H. E., van Dijk, A. I., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from ten state-of-the-art hydrological models, Hydrology and Earth System Sciences, 21, 2881–2903, 2017a.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Gonzalez Miralles, D., Martens, B., and De Roo, A.: MSWEP: 3-hourly 0.25 global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, 2017b.

Bevacqua, E., Maraun, D., Haff, I. H., Widmann, M., and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), Hydrology and Earth System Sciences, 21, 2701, https://doi.org/10.5194/hess-21-2701-2017, 2017.

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., and Widmann, M.: Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change, Science advances, 5, eaaw5531, https://doi.org/10.1126/sciadv.aaw5531, 2019.

Bevacqua, E., Vousdoukas, M., Zappa, G., Hodges, K., Shepherd, T., Maraun, D., Mentaschi, L., and Feyen, L.: Global projections of compound coastal meteorological extremes, https://doi.org/https://doi.org/10.31223/osf.io/4x2u8, 2020.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., et al.: Changing climate both increases and decreases European river floods, Nature, 573, 108–111, 2019.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P.: An introduction to statistical modeling of extreme values, vol. 208, Springer, 2001.

Couasnon, A., Eilander, D., Muis, S., Veldkamp, T. I., Haigh, I. D., Wahl, T., Winsemius, H., and Ward, P. J.: Measuring compound flood potential from river discharge and storm surge extremes at the global scale, Nat. Hazards Earth Syst. Sci., 20, 489–504, https://doi.org/10.5194/nhess-20-489-2020, 2020.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, 2011.

Eilander, D.: Paired time series of daily discharge and storm surge, Zenodo, https://doi.org/https://doi.org/10.5281/zenodo.3258007, https://doi.org/10.5281/zenodo.3258007, 2019.

Ganguli, P. and Merz, B.: extreme coastal Water Levels exacerbate fluvial flood Hazards in northwestern europe, Scientific reports, 9, 1–14, 2019a.

Ganguli, P. and Merz, B.: Trends in Compound Flooding in Northwestern Europe During 1901–2014, Geophysical Research Letters, 46, 10 810–10 820, 2019b.

Heffernan, J., Stephenson, A., and Gilleland, E.: Ismev: an introduction to statistical modeling of extreme values, R package version, 1.41, https://CRAN.R-project.org/package=ismev, 2016.

Hendry, A., Haigh, I. D., Nicholls, R. J., Winter, H., Neal, R., Wahl, T., Joly-Laugel, A., and Darby, S. E.: Assessing the characteristics and drivers of compound flooding events around the UK coast, Hydrology and Earth System Sciences, 23, 3117–3139, https://doi.org/10.5194/hess-23-3117-2019, 2019.

Khanal, S., Ridder, N., Terink, W., Hurk, B. v. d., et al.: Storm surge and extreme river discharge: A compound event analysis using ensemble impact modelling, Frontiers in Earth Science, 7, 224, 2019.

Klerk, W.-J., Winsemius, H., van Verseveld, W., Bakker, A., and Diermanse, F.: The co-incidence of storm surges and extreme discharges within the Rhine–Meuse Delta, Environmental Research Letters, 10, 035 005, 2015.

Kumbier, K., Carvalho, R. C., Vafeidis, A. T., and Woodroffe, C. D.: Investigating compound flooding in an estuary using hydrodynamic modelling: a case study from the Shoalhaven River, Australia, Natural Hazards and Earth System Sciences, 18, 463, 2018.

Manning, C., Widmann, M., Bevacqua, E., Van Loon, A. F., Maraun, D., and Vrac, M.: Soil moisture drought in Europe: a compound event of precipitation and potential evapotranspiration on multiple time scales, Journal of Hydrometeorology, 19, 1255–1271, 2018.

Manning, C., Widmann, M., Bevacqua, E., Van Loon, A. F., Maraun, D., and Vrac, M.: Increased probability of compound long-duration dry and hot events in Europe during summer (1950–2013), Environmental Research Letters, 14, 094 006, 2019.

Martius, O., Pfahl, S., and Chevalier, C.: A global quantification of compound precipitation and wind extremes, Geophysical Research Letters, 43, 7709–7717, 2016.

Mentaschi, L., Vousdoukas, M. I., Voukouvalas, E., Dosio, A., and Feyen, L.: Global changes of extreme coastal wave energy fluxes triggered by intensified teleconnection patterns, Geophysical Research Letters, 44, 2416–2426, 2017.

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C., and Ward, P. J.: A global reanalysis of storm surges and extreme sea levels, Nature communications, 7, 2016.

Olbert, A. I., Comer, J., Nash, S., and Hartnett, M.: High-resolution multi-scale modelling of coastal flooding due to tides, storm surges and rivers inflows. A Cork City example, Coastal Engineering, 121, 278–296, 2017.

Paprotny, D., Vousdoukas, M. I., Morales-Nápoles, O., Jonkman, S. N., and Feyen, L.: Compound flood potential in Europe, Hydrol. Earth Syst. Sci. Discuss, 2018, 1–34, 2018.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., and Erhardt, T.: VineCopula: Statistical inference of vine copulas, R package version, 2.0.5, 2016.

Vahedifard, F., AghaKouchak, A., and Jafari, N. H.: Compound hazards yield Louisiana flood, Science, 353, 1374, 2016.

van den Hurk, B., van Meijgaard, E., de Valk, P., van Heeringen, K.-J., and Gooijer, J.: Analysis of a compounding surge and precipitation event in the Netherlands, Environmental Research Letters, 10, 035 001, https://doi.org/10.1088/1748-9326/10/3/035001, 2015.

Vandenberghe, S., Verhoest, N., Onof, C., and De Baets, B.: A comparative copula-based bivariate frequency analysis of observed and simulated storm events: A case study on Bartlett-Lewis modeled rainfall, Water Resources Research, 47, 2011.

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., and Feyen, L.: Extreme sea levels on the rise along Europe's coasts, Earth's Future, 5, 304–323, 2017.

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Jevrejeva, S., Jackson, L. P., and Feyen, L.: Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard, Nature communications, 9, 2360, 2018.

Wahl, T., Jain, S., Bender, J., Meyers, S. D., and Luther, M. E.: Increasing risk of compound flooding from storm surge and rainfall for major US cities, Nature Climate Change, 5, 1093–1097, 2015.

Ward, P. J., Couasnon, A., Eilander, D., Haigh, I. D., Hendry, A., Muis, S., Veldkamp, T. I., Winsemius, H. C., and Wahl, T.: Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries, Environmental Research Letters, 13, 084 012, 2018.

Wu, W. and Leonard, M.: Impact of ENSO on dependence between extreme rainfall and storm surge, Environmental Research Letters, 14, 124 043, 2019.

Wu, W., McInnes, K., O'grady, J., Hoeke, R., Leonard, M., and Westra, S.: Mapping Dependence Between Extreme Rainfall and Storm Surge, Journal of Geophysical Research: Oceans, 123, 2461–2474, 2018.

Yamazaki, D., O'Loughlin, F., Trigg, M. A., Miller, Z. F., Pavelsky, T. M., and Bates, P. D.: Development of the global width database for large rivers, Water Resources Research, 50, 3467–3480, 2014.

Zscheischler, J., Westra, S., Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., et al.: Future climate risk from compound events, Nature Climate Change, p. 1, 2018.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N., Thiery, W., and Vignotto, E.: A typology of compound weather and climate events, Nature Reviews Earth & Environment (Accepted for publication), 2019.

Answers to comments from anonymous referees 1 and 2

We would like to thank the reviewers for the time spent in reviewing the paper. We found the comments and suggestions very valuable and constructive. We believe that they contributed to improving the manuscript.

Please, find the response to the individual comments below. Note some tables and figures are located at the end of the file (they are relevant for addressing different comments from both reviewers).

Best regards,
Emanuele Bevacqua (on behalf of all authors)

Anonymous Referee #1

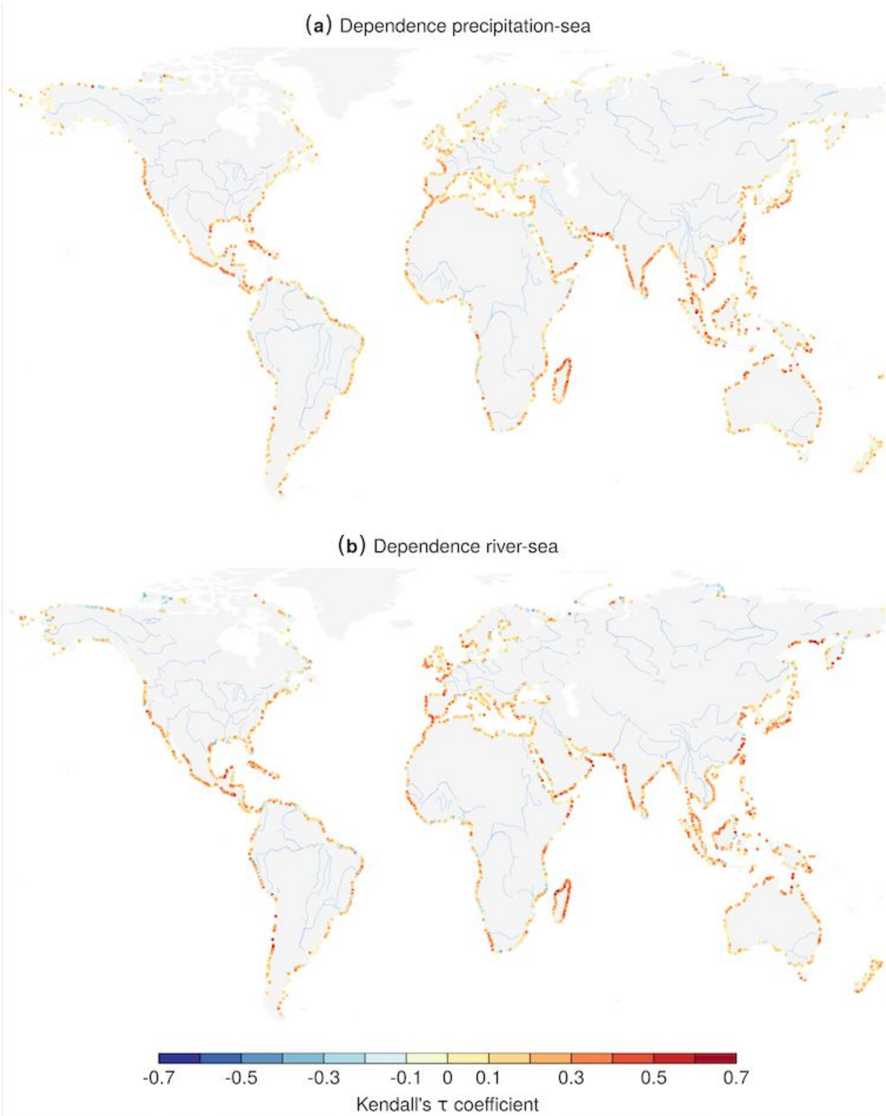
> *(1) This manuscript evaluates the risk of coastal compound flooding at the global scale by using different combinations of drivers (surge vs precip and surge vs discharge) and assess the differences in the results. The analysis is purely based on modelled data. I think the exercise is interesting and provides some important insights that can guide future large-scale assessments of compound flooding risk. The paper is very well written and pleasant to read. I provide some critical comments below which should be addresses before publishing the manuscript.*

We would like to thank the reviewer for the very positive feedback.

> *(2) One general comment is that a CF analysis usually starts with assessing dependence between the drivers of interest, in the context of copulas often based on Kendall's rank correlation. This step is completely skipped in the analysis, but a simple comparison of the rank correlation between variable pairs would already provide useful information, before moving on to the more complex statistical analysis. It would also show where correlation is small/negligible and a complex joint probability analysis possibly not warranted.*

One possibility would be to analyse the dependence between the full distribution of the data, i.e. without focussing on the tails. We think that this might be misleading because, in principle, data might show a poor correlation in the bulk of the distribution but high tail dependence. One could also look at the dependence between the selected "extreme-pairs" used for the copula fitting. Response Fig. 1 shows the dependencies of the pairs used for the fit of the copulas (we show Kendall's tau correlation associated with the copulas fitted to the pairs). A visual inspection of the maps shows that some of the spatial pattern of high dependencies resembles that of low return periods; higher dependencies are found, e.g., along parts of Madagascar and India. However, a low dependence in the tail may also hide a low-return period, if the interarrival time between the selected pairs is low (and vice versa) (see return period expression). Also, the dependence in locations with a different number of selected-pairs may convey different physical information. For these reasons, we have preferred to directly analyse the CF return periods at all grid points, which in this context provide information on the overall dependence of the pairs (including the behaviour in the tail of the bivariate distribution).

To make things more clear, we have added Response Fig. 1 in the appendix. We refer to this figure in the methods, where we also write*: "Note that, as the return period is obtained as a combination of the average elapsing time \mu and the parametric probability density function of the data in the tail, an exact correspondence between the dependence of the copula and the return period is not expected."* In addition, Response Fig. 1 is also relevant in relation to the next comment of the referee 1 (see below).



Response Fig. 1. Kendall's tau correlation associated with the copulas fitted to the selected pairs of (a) precipitation and sea level, and (b) river discharge and sea level.

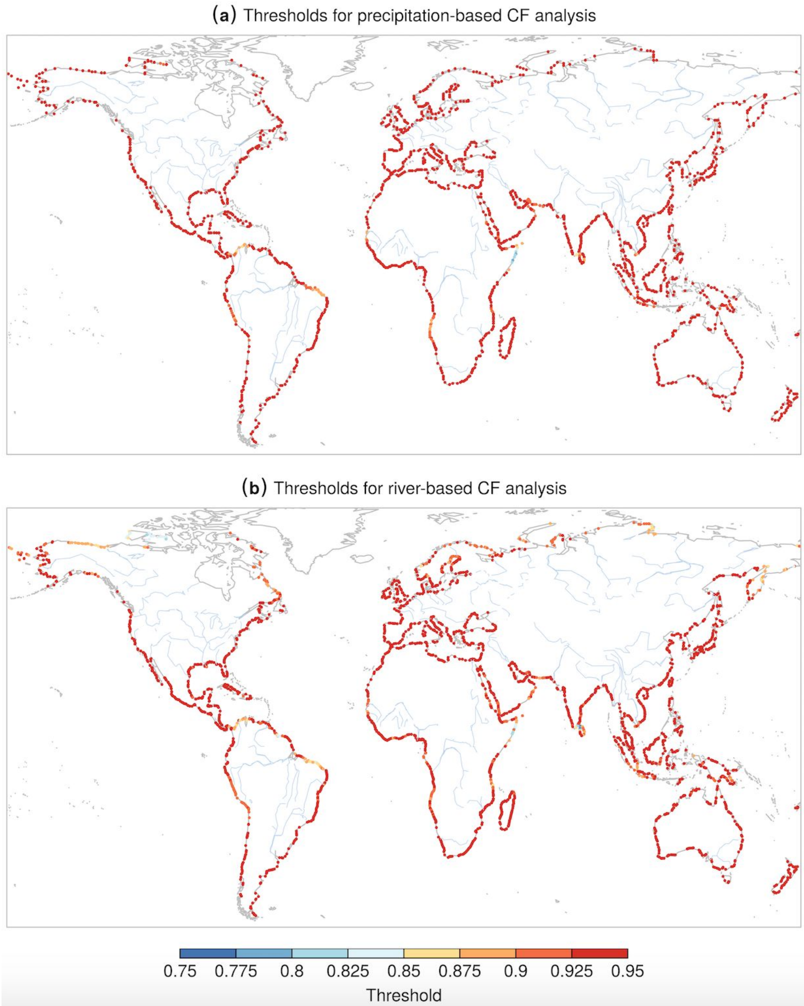
> *(3) L. 94 The threshold of having at least 20 values seems quite low and I suspect the rank correlation that feeds into the copula analysis to be quite sensitive to*

individual extreme event combinations when the number of pairs is so low. This alone could lead to large differences between the CF estimates for the different variable pairs. It would be interesting to see if there is a relationship between the differences in CF estimates and the data availability, i.e., CF ratio vs sample size. [...] What is the lowest threshold that has to be used in order to get to 20 events?*

*The part of the comment in brackets is addressed later

Thanks for the comment. Employing extreme value theory leads to the advantage of focusing on the tail of the distribution, but also to the disadvantage of having a limited amount of data. In this context, we adopted 20 pairs as a tradeoff between having not too few pairs and having a relatively high threshold. Furthermore, the results are robust with respect to changes in the value N=20. As a first check for the sensitivity of the rank correlation to the number of pairs, please see Response Fig. 1, showing the Kendall's tau of the pairs. The maps indicate that the rank correlations vary pretty smoothly in space, supporting the robustness of the assessment. (Note that, in both maps, some small scale spatial variation in the dependencies may also arise from physical processes. Different catchment characteristics in neighbouring rivers may lead to spatial variations in the river-based dependence; while, to a certain extent, local precipitative events may lead to spatial variations in the precipitation-based dependence.)

Related to "*What is the lowest threshold that has to be used in order to get to 20 events?",* the thresholds used for the fit are relatively high, as shown in Response Fig. 2. See also Response Table 1, showing some specific values of the distribution of the thresholds used in the two analyses, including the minimum values asked explicitly by the referee (based on N=20 pairs). The number N=20 is a compromise between having a large number of pairs, and avoiding obtaining too low thresholds where the values would not be extreme. In fact, to choose this number we had checked the same statistics for N=30 (table 2) and N=40 (table 3). The number N=20 was chosen as it allows for using a threshold above the 90th percentile in at least 90% of locations in both analyses.



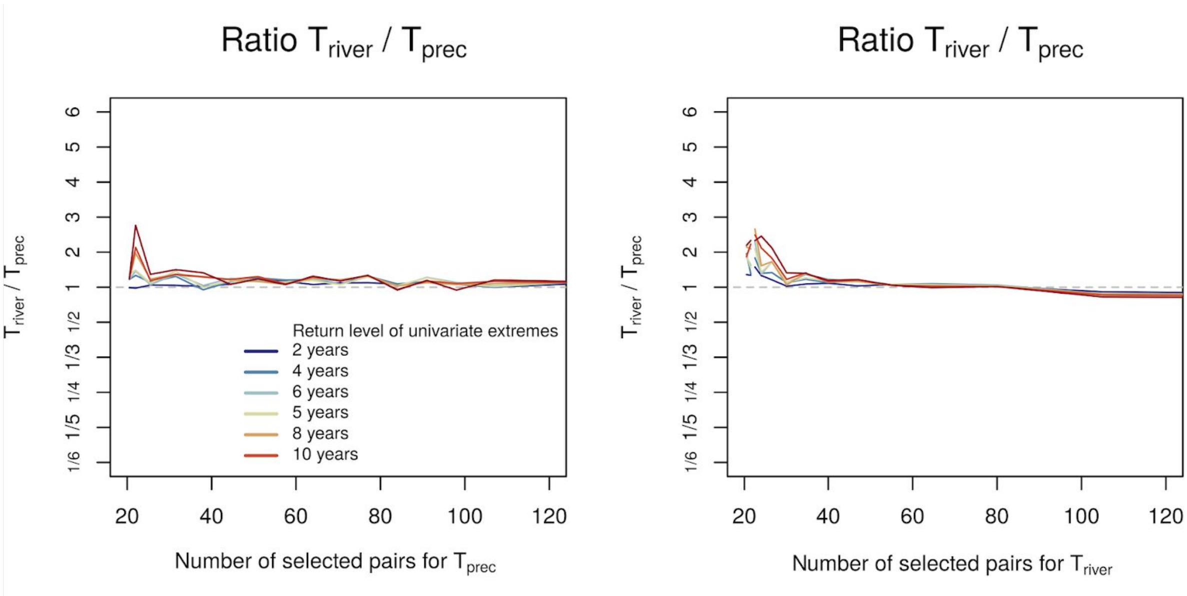
Response Fig. 2. Thresholds employed for the (a) precipitation and (b) river based assessments.

Based on the suggestion of the referee, we show the CF ratio conditioned on the sample size of the selected pairs (Response Fig. 3). Given that the number of selected pairs depends on the analysis (river- or precipitation-based), we show two plots. No clear relationship between the median ratio Triver/Tprec (thick lines) and the sample size appears from Fig. 3. However, there is a weak tendency towards slightly higher ratios for smaller sample size, therefore we did carry some additional tests to make sure that the results are not affected by a low number in selected pairs.

We carried out several tests during the development of the study. These were important to support the choice of N=20 pairs and are relevant in this context also to answer the concerns of the reviewer regarding the effect of a potentially too low value N. The tests highlighted that the return periods are robust, i.e. insensitive to reasonable changes in this value N. This is shown in the Response Fig. 8 at the end of the file for return periods associated with return levels of 2 and 5 years, and for N=20 (used in the manuscript) and N=40. Therefore, no differences in the two assessments are expected because of the choice of N=20. Given the robustness of the results for changes in N, we did choose N=20 that, as discussed above, is a compromise between having a large number of pairs and avoiding obtaining too low thresholds where the selected values would not be extreme.

We now discuss the topic of the thresholds explicitly in the paper, which is actually very important to avoid any potential concerns to the reader:

"If the defined thresholds result in a small group of selected pairs, we lower the 95th percentile selection threshold to guarantee having at least 20 pairs. The choice of 20 pairs is a *tradeoff between having a sufficient amount of selected pairs and* employing a reasonably high threshold *for the fit of the parametric distribution in the tail. Furthermore, the return periods are largely insensitive to reasonable changes in the threshold (results are similar based on 20, 30, and 40 pairs; not shown). The selection thresholds are generally high: 75% of the locations have a selected-threshold larger or equal to 0.95 and 0.94 for the precipitation- and river-based analysis, respectively. And 95% (99%) of the locations have a selected-threshold above 0.93 (0.885) and 0.89 (0.85) for the precipitation- and river-based analysis, respectively "*



Response Fig. 3. Ratio Triv/Tprec against the number of selected pairs. (a) Ratio between the CF return periods based on river discharge (Triver) and on precipitation (Tprec), as a function of the number of selected pairs employed to fit the bivariate parametric distribution for the precipitation based assessment (and of the return levels used to define the CF univariate extremes; see legend in panel (a)). (b) The same as (a), but the ratio is conditioned on the number of selected pairs employed to fit the bivariate parametric distribution for the river based assessment. To obtain the plot, locations are ranked based on the number of selected pairs, and then divided into groups, each containing the same number of locations. For each group, we then compute and show the median ratio. (Note that we employ the same range of the y-axis (1/6, 6) as that of Fig. 3d in the original manuscript, which allows for direct comparison of the figures).

(4) I was also surprised by the short decluster time of only 3 days, given that discharge data is involved in the analysis; the authors actually discuss the aspect of discharge events often lasting longer later on in the manuscript. In many cases discharge can exceed a high threshold for several weeks or even months, and the way the analysis is setup here multiple extreme values are sampled from these events. On the one hand, of course, every time a surge occurs during the high discharge event there could be compound flooding, but on the other hand basic extreme value theory assumptions are violated. I am not saying the entire analysis has to be changed, but would like to hear the authors opinion on this issue, and it might also be worth touching on in the manuscript.

We agree with the referee that "In many cases discharge can exceed a high threshold for several weeks or even months, and the way the analysis is set up here multiple extreme values are sampled from these events." and therefore we agree that this is a point deserving an explicit discussion within the manuscript. The reason for setting up the analysis in this way is exactly that mentioned by the referee, i.e. "every time a surge occurs during the high discharge event there could be compound flooding". Therefore, considering only one of the several storm surges that occur during a period of high river flow would lead, in practice, to an underestimation of the CF potential. (This is clear when thinking, e.g., about New Orleans (Louisiana, US) that in 2019 experienced extremely high water discharge from the Mississippi River from March to July. Any storm surge in that period could have led to compound flood potential, hence they all should be taken into account in an assessment of the hazard.) Therefore, we believe that it is important to follow a pragmatic approach and adopt this choice of considering all the storm surge events. We discuss this topic in the Methods when introducing the declustering, where we add:

"While this choice has the drawback of not fully respecting the assumptions of independent realisations of the extreme events, which is necessary to apply extreme values theory in its generic form, it allows considering multiple storm surges that may occur during a sustained period of high river discharge and that could lead to multiple compound floods."

(5) L. 100 Is the same copula always used when the different variable pairs are analyzed (i.e., get the best fit copula for, e.g., surge vs precip and force the same copula to be used for surge vs discharge) or can it change? If the copula is free to change, it might be interesting to test how the results look like wen the same copula is forced (as long as it passes the goodness of fit test(s)).

First, we observe that the dependence between the variables is captured within the analysis both by the interarrival time between select pairs (that is smaller for higher-dependent data) and by the copula itself. Both, naturally, contribute to the differences between the two assessments.

Regarding the specific question, we observe that for any given location the physical processes captured by the two assessments can differ, therefore there is no physical justification for employing the same copula. In particular, it should be also considered that there are no available tools for defining which is the “right” copula structure and we do not have prior physical knowledge of the true structure of dependence at the different locations. Therefore, we have employed the same criterion for selecting the copula in both individual assessments and allowed the copulas to be different in the two assessments. (We have of course applied statistical tests to investigate whether the fitted copulas could be rejected or not, and the results indicate that the chosen copulas are consistent with the data.) For the reasons above, we think that an interpretation of the proposed analysis would not be straightforward.

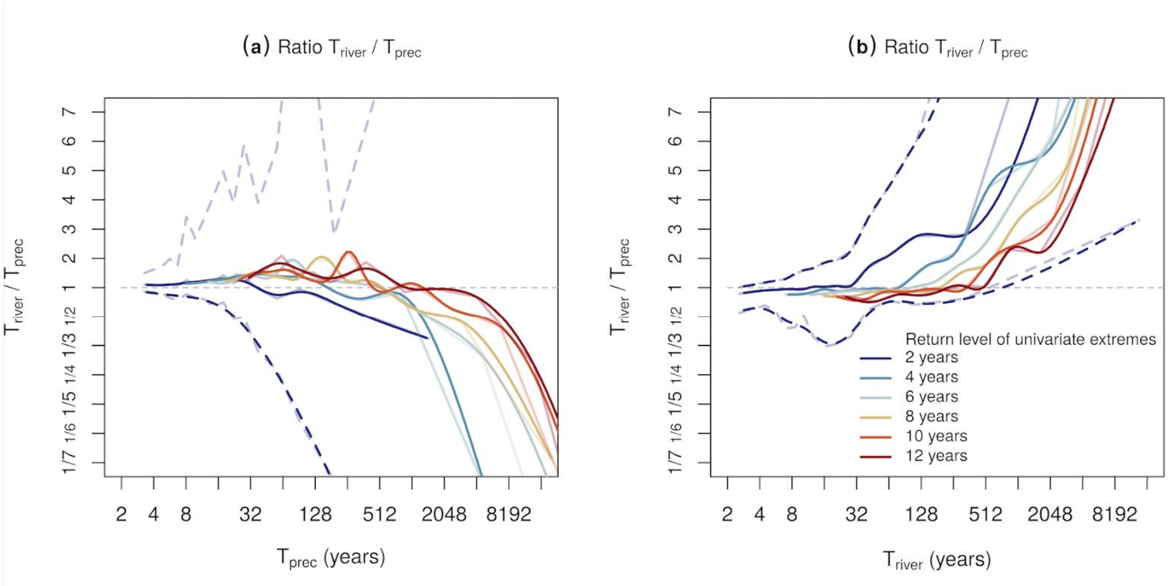
Based on the question of the referee, we have added a sentence in the manuscript where we state explicitly, to avoid misunderstandings, that different copulas are allowed at the same location: “*In general, the physical processes captured by the two assessments can differ (even at the same location), therefore we allow for the selection of different copulas in the two assessments.”*

> *(6) When displaying the results it would be interesting to see the relationship between CF ratio and absolute CF risk (in relation, for example, to the independence assumption), i.e., are differences between the CF risk relatively larger/smaller in areas where the joint return periods are closer/more different to the independence assumption.*

We think that investigating whether the differences between the two return periods is more pronounced for high return periods (of Tprec or Triver) is a very valuable suggestion. The relationship between the CF ratio and the absolute CF return periods is somehow visible from Fig. 1 and 2, however, we had not considered this explicitly. We show the relationship in Response Fig. 4 (added to the new version of the manuscript). The figure reveals that there is a tendency towards higher differences in the two assessments in locations where either or both Tprec and Triver are high. The differences tend to be small at locations with small return periods.

We explain the behaviour seen in Response Fig. 4a based on the following reasoning (similar reasoning applies for Response Fig. 4b). First it should be considered that (1) Tprec and Triver are correlated (Fig.1 and Fig. 3a) and that (2) large return periods are highly uncertain.
The combination of (1) and (2) provides insight on why:

- the ratio Triver/Tprec conditioned to locations characterized by low Tprec is about equal to 1 (Response Fig. 4a). Locations where Tprec is low are characterized, on average, by a low Triver (because of 1). Low return periods have relatively low uncertainties, therefore the selected Triver (associated with low Tprec) is on average similar to Tprec. This ultimately leads to a ratio about equal to 1.
- Similarly, the ratio Triver/Tprec conditioned to locations characterized by high Tprec is low (Response Fig. 4a). Locations where Tprec is high are characterized, on average, by high Triver (because of 1). However, while Tprec is very high and fixed due to the conditioning itself, the selected values of Triver are largely variable because of (2). As a result, the average value of Triver is smaller than the Tprec (causing low ratio).



Response Fig. 4. (a) Ratio between the CF return periods based on river discharge (T_{river}) and on precipitation (T_{prec}), as a function of T_{prec} (and of the return levels used to define the CF univariate extremes; see legend in panel (b)). The dashed line is the centered 68\% (16th-84th percentiles) confidence interval of the ratio based on the 2-year return levels. Tick lines are obtained through regressing the investigated statistical values to the natural logarithm of the mean return period of the bins via a spline function. (b) The same as (a), but the ratio is conditioned on T_{river}. As in Fig. 3d of the original manuscript, a non linear Y-axis for values below 1 is employed such that the specular cases, e.g., ratio r=2 ($T_{river}=2 \cdot T_{prec}$) and r=1/2 ($T_{prec}=2 \cdot T_{river}$) (see Fig. 3b) appear symmetric with respect to r=1 ($T_{river}=T_{prec}$).

Fig. Response Fig. 4 provides insights also on whether the "differences between the CF risk is relatively larger/smaller in areas where the joint return periods are closer/more different to the independence assumption". In fact, the differences between the CF assessments (the divergence from 1) tend to occur - on average - at higher return periods (x-axis) for the assessment based on higher extreme values (colors towards the red). That is, the ratio diverges from 1 while the return period approaches values consistent with independent drivers (that are higher for return periods associated with extremes defined as higher return levels).

Overall, we believe that the suggested investigation provides also some general and interesting insights of practical interest. In fact, it implies that practically the discrepancies between the approaches are large, especially where the actual CF potential tends to be low (where CF drivers approach independence), i.e. where an analysis of compound flooding is less relevant. Hence, the precipitation-based analysis is similar to the river-based where it is actually more relevant to analyse the CF hazard, which strengthens some of the conclusions of the work.

When presenting the results, we write: *"We find that there is a tendency towards higher differences in the two assessments at locations where either or both Tprec and Triver are* *high (i.e. where T approaches the value expected under independence of the CF drivers) (Fig. A6). (This appears consistent with the high uncertainty associated with large CF return periods.) Such a finding has relevant implications, as it indicates that the two assessments tend to be similar, on average, where assessing the actual CF is more important, i.e. where there is a relatively high CF potential (Fig. A6)."*

In the abstract and conclusions, we added: *"on average, the deviations between the two assessments are smaller in regions where assessing the actual CF is more relevant, i.e. where the CF potential is high."*

Anonymous Referee #2

Bevacqua et al., present a global scale analysis of compound flood (CF) potential by comparing results that estimate the probability of CF using precipitation and storm surge water levels (surge and waves) and discharge and storm surge water levels (surge and waves). This is an important topic for CF research, as a variety of results have been published that use either discharge or precipitation and to my knowledge, this is the first study that tests how the two variables compare. This paper was well written and easy to read. While this is a valuable addition to the literature, I'd like to see the authors more thoroughly address that differences between results are not due to their analysis techniques before being accepted for publication.

We would like to thank the reviewer for the very positive feedback, highlighting the timeliness of our work.

Specific Comments: The manuscript could address whether results are dependent on the chosen analysis techniques more thoroughly. I'm broadly interested in if the patterns in the return periods and the Triver/Tprec ratio are related to local characteristics of the datasets as the authors suggest in their Results and discussion section, or the analysis that was undertaken, particularly concerning 1) dependency between variables, 2) threshold selection and 3) goodness of fit.

1) Most CF assessments begin by assessing the dependence of the given parameter space. A comparison of the dependence between precipitation and surge level and discharge and surge level may help to explain patterns in the differences in the results. Are there large differences in the dependence between surge vs precip and surge vs river extremes?

Given that the extremes are defined based on return level values, the present analysis of potential compound flooding is directly linked to the dependence of the variables. Therefore, yes, the results are well explained by differences in the dependence. We now add a sentence in the methods: *"Overall, given the definition of the extremes based on α-year return levels, the bivariate return period is inherently linked to the dependence of the pairs in the tail of the distribution."*

Given that this question is similar to that asked by referee 1, please see the answer to comment (2) of referee 1 above.

2) The authors select pairs of data that are larger than the individual variable's 95 %ile threshold. If less than 20 pairs, the authors decrease the threshold to include more pairs of joint variables. How was the number of pairs, 20, chosen? This seems like a fairly low amount of joint-events to base an analysis on (less than one a year!). Furthermore, how many locations was the threshold decreased, and what's the lowest threshold that was used? Can these variables still be considered "extreme?" It would be interesting to know how variable the number of events analyzed per location was in this analysis.

First of all, it could be relevant to highlight that all the variables are used to estimate the return period, not only the selected pairs. As shown in equation 1 in the main text, the return period is a function of the interarrival time between selected pairs and the parametric probability density function modelling the selected pairs. That is, this is a semi-parametric return period (line 90 of the original manuscript), as it is always the case for the corresponding univariate case of return periods based on peaks over threshold.

We agree that the number of N=20 pairs is somewhat arbitrary. However, it was chosen after carrying some tests (both during this and previous studies) for the bivariate return periods. About the variability of the number of selected pairs, there is a tendency towards having fewer pairs at locations where CF is unlikely. This is consistent with the fact that there are fewer concurrences of extremes there. The median number of selected pairs for the river-based analysis is 30 (interquartile range (22,57)), and for the precipitation-based analysis is 58 (interquartile range (30,86)).

The tests that we carried out to choose N=20 as threshold showed that the return periods are robust with respect to reasonable changes in this value N. This is shown in the Response Fig. 8 at the end of the file for return periods associated with return levels of 2 and 5 years, and for N=20 (as in the manuscript) and N=40.

Furthermore, to answer other specific questions, please see Table 1 at the end of the file. It shows some specific values of the distribution of the thresholds used in the two analyses (based on N=20 pairs). The number N=20 is a compromise between having a large number of pairs, and avoiding obtaining too low thresholds whereas the values would not be extreme anymore. In fact, to choose this number we had checked the same statistics for N=30 (table 2) and N=40 (table 3). The number N=20 was chosen as it allows for using a threshold above the 90th percentile in at least 90% of locations in both analyses. (For example, using N=40, the lowest threshold would be 0.68, i.e. we would not be in the tail of the distribution anymore.) This is important to stay within the framework of extreme values, as also the referee points out asking “Can these variables still be considered “extreme?”. Hence, based on the way we chose N, the variables can indeed be reasonably considered as extremes.

To explicitly address this topic, we have extended the method section:

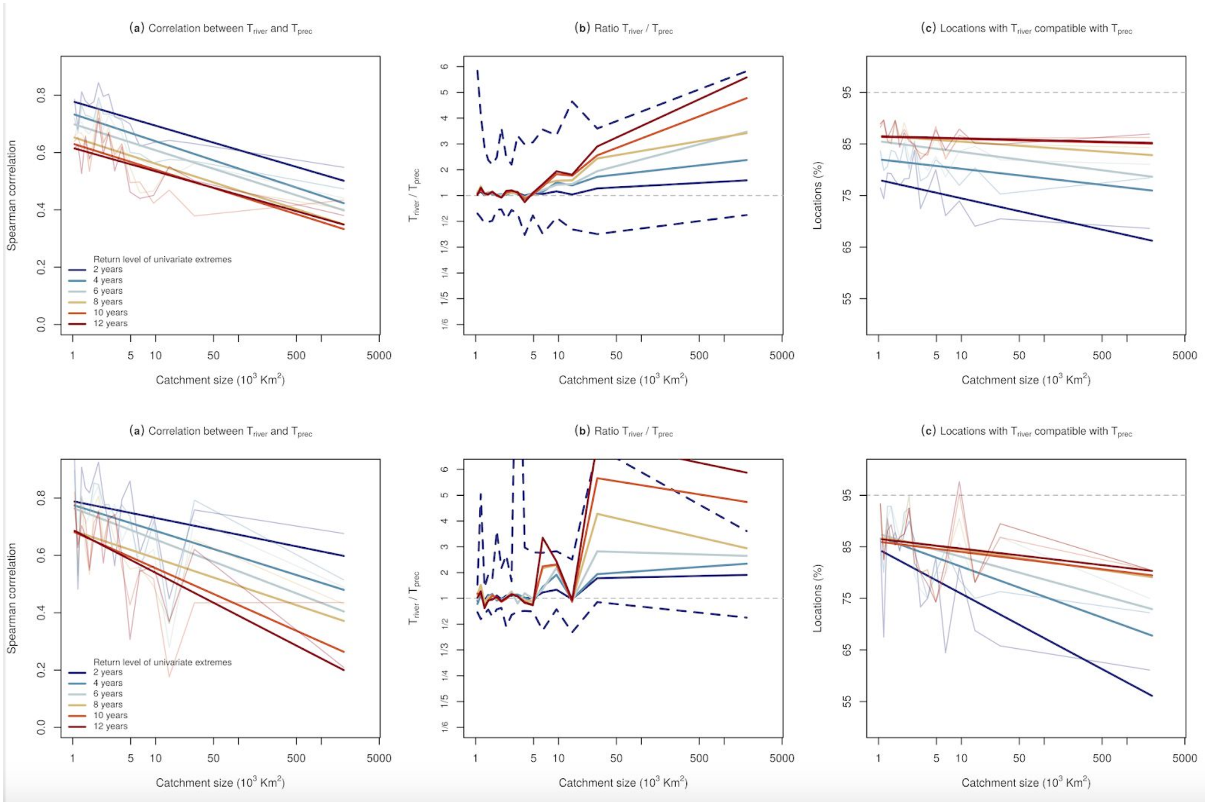
“If the defined thresholds result in a small group of selected pairs, we lower the 95th percentile selection threshold to guarantee having at least 20 pairs. The choice of 20 pairs is a *tradeoff between having a sufficient amount of selected pairs and* employing a reasonably high threshold *for the fit of the parametric distribution in the tail. Furthermore, the return periods are largely insensitive to reasonable changes in the threshold (results are similar based on 20, 30, and 40 pairs; not shown).* T*he selection thresholds are generally high: 75% of the locations have a selected-threshold larger or equal to 0.95 and 0.94 for the precipitation- and river-based analysis, respectively. And 95% (99%) of the locations have a selected-threshold above 0.93 (0.885) and 0.89 (0.85) for the precipitation- and river-based analysis, respectively ”*

3) Do patterns in statistical compatibility have anything to do with the goodness of fit of the marginal distributions and copulas? There are some big differences in statistically compatibility and ratio in the same regions (e.g., stations next door to one another). Any reason why that might be?

In the following, the main concepts are highlighted in bold.

As stated in the methods, the goodness of fit of the marginal distributions and of the copulas were tested, i.e. three distributions (marginals and copula) per assessment were tested (for a total of six distributions tested per location given the two assessments). For a given distribution, i.e. a marginal or a copula distribution, less than about 5% (depending on the distribution) of the p-values are below 0.05, which is an acceptable range as 5% is the value expected under the hypothesis that the data are distributed according to the fitted distribution (Zscheischler et al., 2017). (Note that there is no evident spatial pattern of the locations with p-values below 0.05.)

Furthermore, **we did not find a relationship between the p-values and the statistics/results**. This is shown in *Response Fig. 5*, which is the same as Fig. 3c-d of the manuscript, but obtained as it follows. The first row shows the panel obtained when focussing the analysis only on locations where all the distributions cannot be rejected at 95% confidence level; the second focuses on locations where at least one distribution can be rejected. Naturally, as the latter is based on a lower number of locations, the associated results are more noisy. However, the main conclusions obtained from the two analyses are the same. (Because of this result and of the reasoning of the previous paragraph, we have considered all the locations in the analysis.) Similarly, we also visually inspected (not shown) the map of the ratio Triver/Tprec based only (i) on locations where all the distributions cannot be rejected at 95% confidence level and (ii) only locations where at least one distribution can be rejected. Consistently with the above, in both maps neighbor locations can have large differences in the ratio (ratios>>1 and ratio<<1). Therefore, goodness of fit seems to not be related to the large differences seen in neighborn locations.

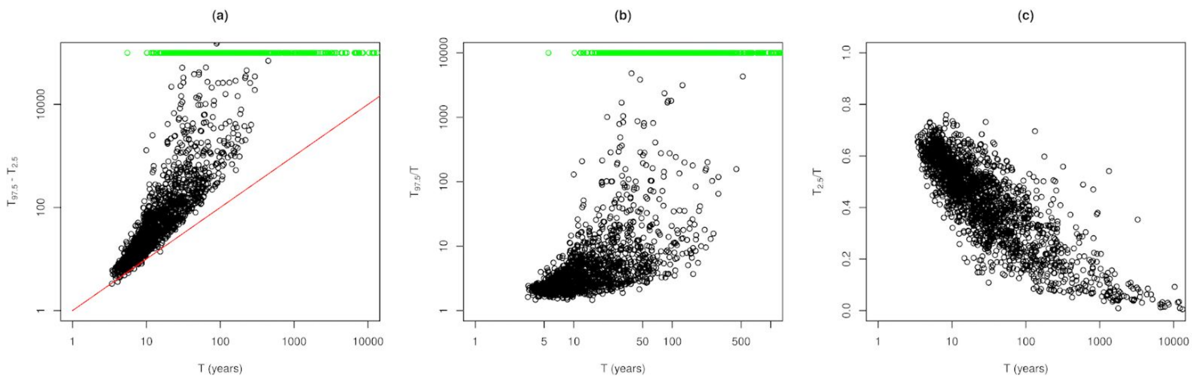


Response Fig. 5. *As Fig. 3c-e of the original manuscript, but the first line shows the results obtained for locations where all the distributions involved in the return period computation "pass" the test, and in the second line the location where at least one distribution does not "pass" the test (see text). In panel b, we show the obtained statistical values per bin with the thick line. In a and c, as in the paper, thick lines are obtained through linearly regressing the obtained statistical values per bin (shown with thin lines) to the natural logarithm of the mean size of the catchment bins.*

Some of these differences (in the map of the ratio Triver/Tprec) between neighboring locations may be explained by differences in basin characteristics. However, we think that the main source of these differences could be return period uncertainties. Indeed, neighborn locations having large differences in the ratio appear frequently in locations where return periods tend to be high, for example in the tropics (Fig. 2). Large return periods are subject to large uncertainties both because of sampling uncertainty and because of model uncertainties (e.g. bivariate model fitting and definition of the return levels). This is indicated by Response Fig. 6a where the 95% confidence interval of Tprec at different locations is shown as a function of the return period Tprec. As a result of the high uncertainty associated with high return periods, in areas where CF is unlikely (e.g., in small catchments of the tropics), return periods are particularly uncertain, hence the estimated return periods can be substantially different (positively or negatively) from the expected one (i.e., the "real" one) (this is indicated by Response Fig. 6b-c). As a result, in locations with high return periods, even assuming that the expected return period from the two assessments being similar, substantial discrepancies in the two assessments (leading to high (>>1) or low (<<1) ratios) may arise from return period uncertainties.

This effect provides a plausible explanation, from a statistical point of view, of why it is possible to find ratio>>1 and ratio<<1 in neighborn locations. The effect can also explain differences in compatibility in nighborn locations: these are consistent with the effect itself combined with the fact that the actual expected return periods from the two assessments can be different enough to lead to differences that are not only large but also significant.

We have added a paragraph in the paper: “In areas with a tendency towards high CF return periods, e.g. the tropics, neighbour locations show divergent values in the ratio between the return periods of the two assessments (dark blue and red dot in Fig. 2). Further tests showed that this behaviour is not related to goodness of fit of the bivariate distributions (see discussion in the response to the reviewer 1), rather it appears associated with the large uncertainties of high return periods and potentially with different catchment characteristics.”



Response Fig. 6. (a) Amplitude of the centered 95% confidence interval of T (i.e., difference between the percentiles: T_97.5th - T_2.5th) as a function of the T. Each dot is a different location, and the figures are obtained based on the precipitation based assessment (similar results would be obtained based on river discharge). Panel b and c indicate that very large (>>1) or small (<<1) values of the ratio between the estimated T and the actual T can occur for large T as a result of uncertainties. (b) Ratio $T_{97.5} / T as a function of T. (c) Ratio T_2.5 / T as a function of T. In (a) and (b), green dots mark locations for which the value on the y-axis is infinite. The 95% confidence interval of T is obtained based on resampling procedure (see Methods).

> *I think it would be helpful to bring up some more information about the catchments studied earlier. What’s the smallest catchment considered? What’s the largest? Are the results dependent on how the catchment values are binned?*

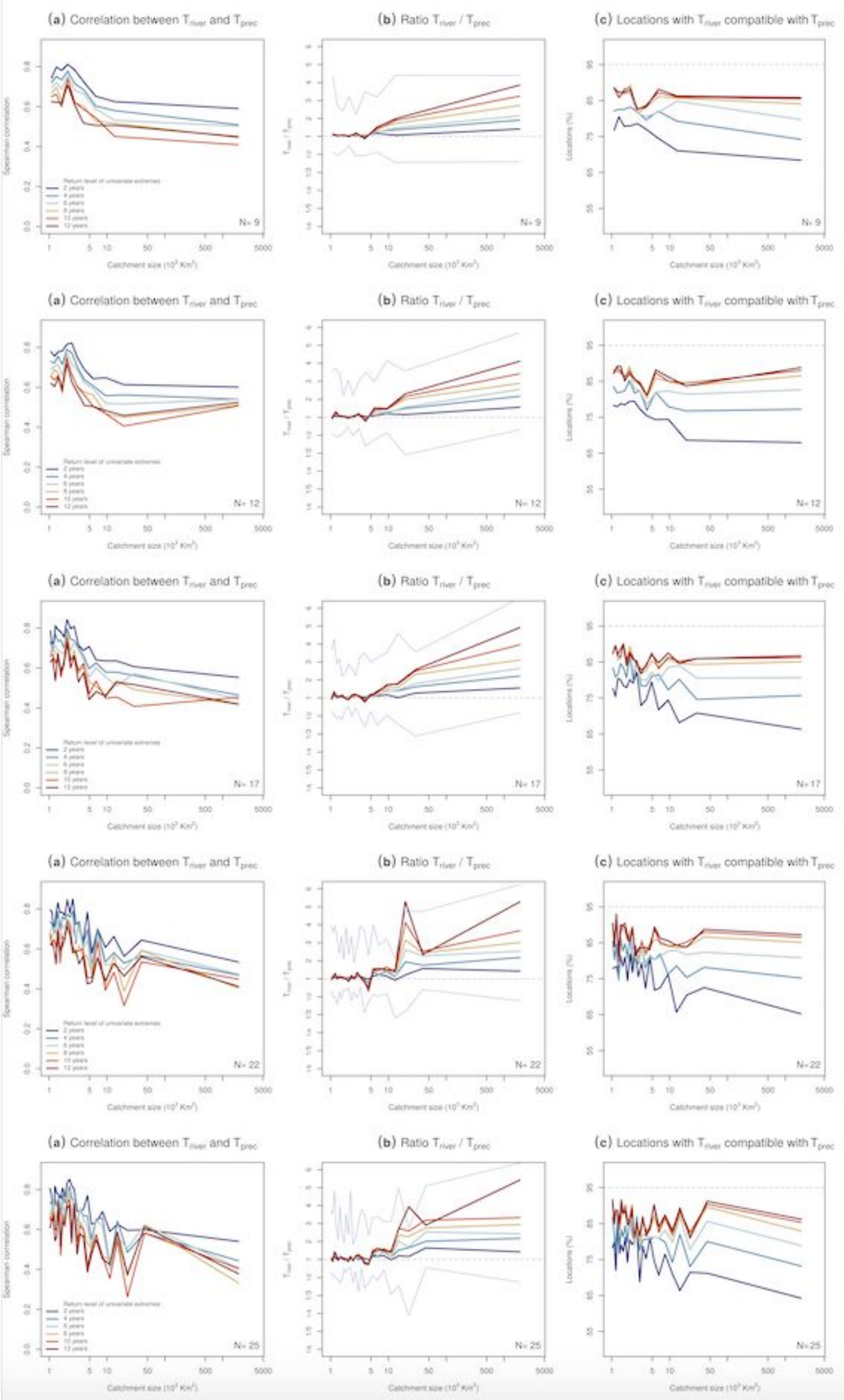
We agree. We have now added information on the catchment at the beginning and end of the data section:

- *“We consider river discharge daily maxima from a publicly available global dataset (Eilander, 60 2019; Couasnon et al., 2019), which includes coastal catchments larger than 1,000 km^{2}.”*
- *“We analyse CF only around river mouth locations whose nearest precipitation and storm surge grid points lie within a distance of 75 km (Couasnon et al., 2019). This results in considering locations at river mouths of catchments with size in between*

about 1,000 and 3,690,000 Km^{2} (95\% having size smaller than 50,000 Km^{2}; Fig. 3f).”

About the binning: Overall, the results of the binning procedure are meant to provide qualitative information, as we also state in the paper: “We qualitatively investigate how the two assessments compare for different classes of catchment size.” (line 112), and adopted this binning procedure as it “provides equally robust statistics for each bin” (line 115).

In the manuscript we state that “we rank the rivers based on their catchment size and divide them into groups having the same sample size; for each group we compute different statistics to compare the two assessments”. It is possible to bin rivers in larger or smaller numbers of groups, corresponding to less or more populated groups, respectively. A too less populated group would lead to more noisy results, on the contrary a too small number of groups (more populated) would not allow for satisfactorily exploring the variability of the statistics as a function of the catchment size. The chosen number of groups is therefore a trade off between these two situations above. **While the results are slightly different when changing the number of groups, the conclusions are unchanged (as mentioned in the original version of the manuscript), e.g., see Response Fig. 7. In the paper, we employ N=17.**



Response Fig. 7. *As Fig. 3c-e of the original manuscript, but employing a different number of bins (shown at the bottom-right of each panel).* Compared to the images of the paper, we do show only the rough statistics (without regressing any spline or line to them, which allows for better appreciating the differences between the graphs).

Finally, on Page 2, line 53-54, the authors state that the study "aims to assess whether a precipitation based CF assessment can be used as a surrogate for potential CF in estuaries." But I feel like the authors never come back to answering this question. For example, most places that are statistically incompatible are where Tprecip is much larger than Triver. Does this analysis suggest then that precipitation can't be used here, or it should be used in these cases?

We thank the reviewer for this comment. In locations where the two return periods are incompatible, one should not employ precipitation for assessing the river based return period. The main aim of the study is not to address whether the two return assessments are interchangeable at the specific locations, rather when carrying out large-scale assessments. However, we agree that also indications for local assessments can be provided.

Although the discussion in the introduction was based on the large-scale CF assessment, we see that in the specific text quoted by the referee, and more specifically in the final paragraph of the introduction, there is no reference to the "large-scale" nature of the assessment. This could indeed be misleading for the reader. Therefore, we have now added some words/sentences in the text (shown in below). In the introduction:

"Given the scarcity and heterogeneous distribution of in situ data (Ward et al., 2018; Couasnon et al., 2019; Wu et al., 2018), scientists have started to employ model data - of river, storm surge, and precipitation - to assess the ***large-scale*** *potential CF hazard (Ward et al., 2018; Bevacqua et al., 2019a; Wu et al., 2018; Couasnon et al., 2019; Wu et al., 2018; Paprotny et al., 2018; Bevacqua et al., 2019b). Against the foregoing background, the present study aims to assess whether a precipitation based* ***large-scale*** *CF assessment can be used as a surrogate for potential CF in estuaries at the* ***large-scale****. To that end we use coherent global model datasets of storms surges (including wave effects) (Vousdoukas et al., 2018), precipitation (Beck et al., 2017b), and river discharge (Couasnon et al., 2019; Eilander, 2019) and conduct a first global comparison of the results obtained through the two approaches, keeping all the other methodological aspects identical."*

Within this context, we feel like we address the question related to assessing "whether a precipitation based **large-scale** CF assessment can be used as a surrogate for potential CF in estuaries at the **large-scale**". In the conclusions, we explicitly refer to the large-scale nature of the assessment and based on the analyses presented, we state: "This study indicates that for these large-scale assessments, a precipitation-based CF analysis can provide satisfactory information on the CF potential in estuaries of small and medium size rivers (catchment smaller than about 5-10,000 Km2)."

Furthermore, to even more explicitly address the comment of the referee about specific locations, we have also modified the one of the last sentences of the paper:

"Naturally, employing river discharge data should always be preferred to using precipitation when studying both the large- and local-scale CF in estuaries, when data are available,

especially in areas where we detected large differences between the two approaches*.* *The importance of using river discharge data is even greater in estuaries of long rivers.”*

> *Comments on Specific Lines:*
>
> *Page 1, line 6: In the abstract, the authors state that CF in long river catchments is “more accurately” analyzed using river discharge data. However, there’s no underlying assessment of the accuracy of either of these datasets, and/or how well these locations represent joint variables at known locations. Thus the authors may want to consider a change in word choice here.*

We changed the text: “...CF in long rivers (catchment $\gtrsim$ 5-10,000 Km^2) should be analysed using river discharge data.”

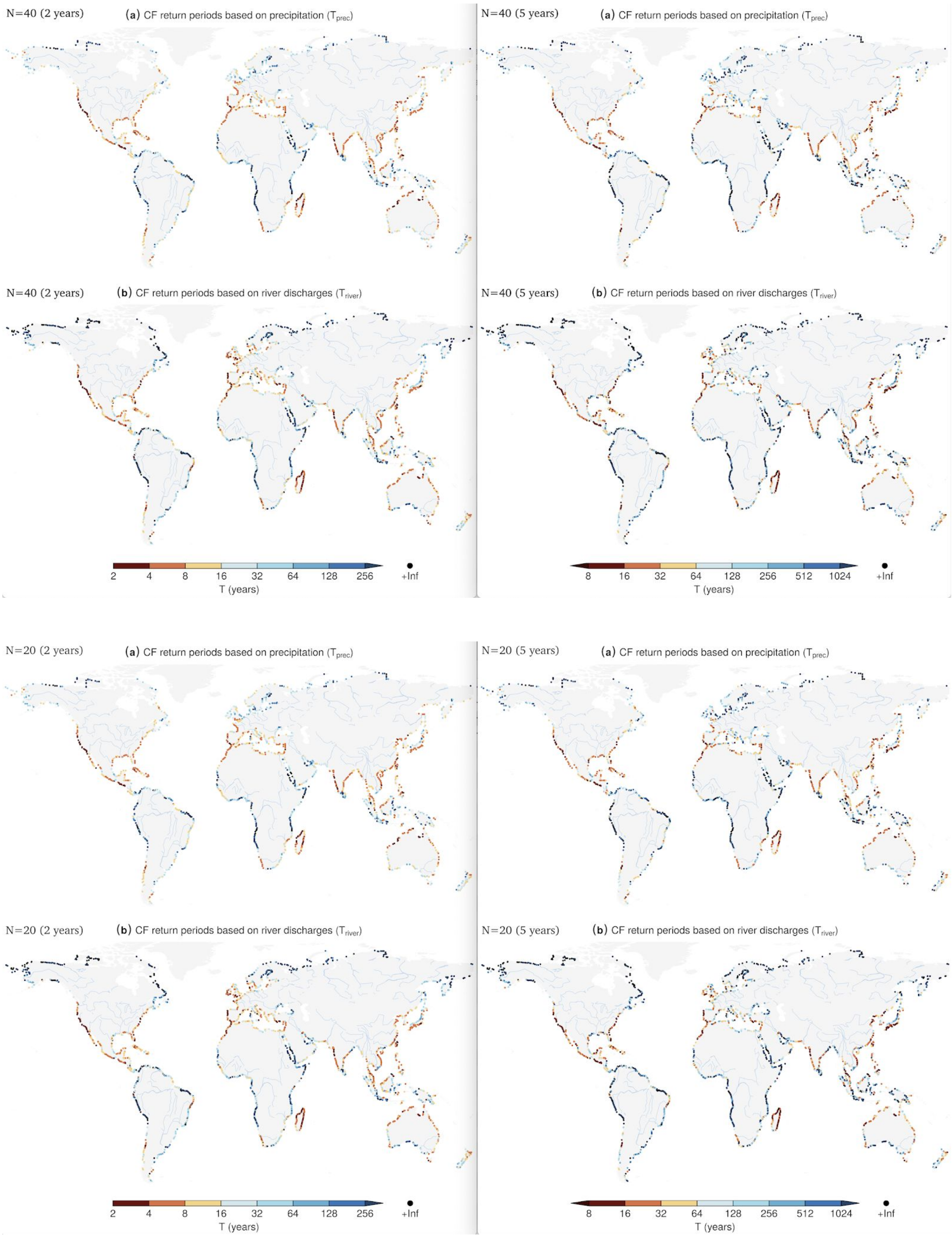
> *Technical Corrections:*
>
> *Page 4, line 108: “centered” is spelled “centred”*
> *Page 10, line 199: “Overall, we find that independently of catchment. . .” should be “independent”*
> *Figure 3: I don’t seem to understand the difference in light versus dark lines in 3d and 3e.*
> *All maps could have lat/lon grid lines.*

Thanks, we have adopted these changes, apart from the lat/lon grid lines that - in our opinion - are not crucial in the context of the paper. However, we would be certainly happy to add them at a later stage if also the editor shares the same opinion as the reviewer.

References

Zscheischler J, Seneviratne SI. Dependence of drivers affects risks associated with compound events. Science advances. 2017 Jun 1;3(6):e1700263.

Material for both reviewers



Response Fig. 8. Comparison between the assessments based on a different number of minimum required selected pairs for the fit of the bivariate parametric distribution. Top 4 panels based on N=40 pairs; Bottom 4 panels based on N=20 pairs. We show the return periods based on extremes defined as 2- and 5-year return levels (as in the manuscript) on the left and right column, respectively.

Nmin=20		
	Selected-threshold of the precipitation-based analysis	Selected-threshold of the river-based analysis
Minimum value	0.785	0.775
1st percentile	0.885	0.84
10th percentile	0.935	0.91
25th percentile	0.95	0.94
50th percentile	0.95	0.95

Response Table 1

Nmin=30		
	Selected-threshold of the precipitation-based analysis	Selected-threshold of the river-based analysis
Minimum value	0.765	0.745
1st percentile	0.86385	0.815
10th percentile	0.92	0.895
25th percentile	0.95	0.925
50th percentile	0.95	0.95

Response Table 2

Nmin=40		
	Selected-threshold of the precipitation-based analysis	Selected-threshold of the river-based analysis
Minimum value	0.755	0.68
1st percentile	0.845	0.795
10th percentile	0.905	0.88
25th percentile	0.935	0.91
50th percentile	0.95	0.935

Response Table 3