Author Response to Referee #2: Assessing the spatial correlation of potential compound flooding in the United States

Huazhi Li, Robert A. Jane, Dirk Eilander, Alejandra R. Enríquez, Toon Haer, and Philip J. Ward

---

General comment

The study proposes a statistical framework for evaluating the spatial correlation of compound flooding, demonstrating its use for some coastal locations in the USA. While the methodology appears potentially valuable for flood risk assessments in coastal cities, the manuscript currently lacks sufficient detail to fully convey its approach and practical implementation. This limits the reader's ability to understand and apply the proposed framework. Furthermore, the accompanying scripts intended to reproduce the analysis are incomplete.

We would like to thank the reviewer for their time and effort dedicated to reviewing our manuscript. We are pleased to read that the reviewer finds the presented methodology potentially valuable for flood risk assessments in coastal cities.

In response to the detailed and constructive feedback, we have thoroughly revised the manuscript. Specifically, we have:

1. Expanded the description of our methodological framework by adding the necessary mathematical formulations and a schematic illustrating the construction of spatially joint events;
2. Clarified several methodological choices, including the use of daily data and the choice of infilling data gaps; and
3. Acknowledged the key limitation related to ignoring the mixed-population effects.

In addition, we have reviewed the available scripts and uploaded all required functions and scripts to the current repository. Both the dataset and code repositories are now openly accessible to facilitate a better understanding and broader adoption of the proposed approach and developed datasets. The dataset can be retrieved at: https://doi.org/10.5281/zenodo.15728000; and the scripts are available at: https://doi.org/10.5281/zenodo.17464793.

We believe that these revisions have substantially improved the manuscript. Our detailed response to specific comments are provided in the sections below.

Major comments

1. L108: Why did the authors use daily data and not instantaneous (e.g., hourly or 15-minute) data?

Thank you for your comment. When characterising dependence between compound flood drivers, standard extreme-value theory statistical models (e.g. copulas, max-stable processes, and the applied multivariate conditional model) require independent and identically distributed (i.i.d) random variables as input. This is commonly achieved by detrending and de-clustering the raw time series. As extreme sea level and high discharge events often last several days, the typical time window for de-clustering is more than 12 hours. Therefore, daily values would be a better choice for this process compared to instantaneous data. In addition, USGS discharge observations are only available at a daily time scale while the original sea level records are hourly and sometimes 6-min for particular stations.

To maintain consistency between these two datasets, we aggregate the instantaneous sea levels into daily maxima.

We have reflected on this point in the manuscript:

> L118: When characterising dependence, standard extreme-value theory statistical models require that the input datasets consist of independent and identically distributed (i.i.d) variables. To satisfy this assumption, we first detrend the hourly total water level records by removing the long-term mean sea level signal. This is achieved by subtracting the annual mean sea level using a moving window, thereby filtering out the inter-annual to multi-decadal sea level variability (Valle-Levinson et al., 2017). River discharges do not show such long-term variations, and so no detrending is applied to the daily mean records. To prepare for the independence processing and maintain temporal consistency between total water level and river discharge, we further aggregate the hourly sea levels into daily maxima. The independence is then ensured by applying a 5-day de-clustering window (Camus et al., 2021; Maduwantha et al., 2024) with the maximum value centred in each window.

2. L120-139: Instead of filling up data gaps based on information from nearby stations, would have been possible to perform all the analyses only considering time periods and storms unaffected by these gaps? Please discuss pros and cons of the two alternative approaches.

Thank you for raising this good point. As the applied statistical model and other similar models like copulas struggle to process missing values, considering only time periods where data are available at all locations would likely lead to very few overlapping events. Using this small sample may be insufficient to calculate the dependence structure. In contrast, using the complete data with infilled missing values largely increases the number of extreme events but may introduce artificial signals and may unintentionally increase the correlation between these locations.

We have included this discussion in the manuscript:

> L127: Gauge observation records often suffer from data gaps and may preclude a robust statistical dependence analysis between flood drivers. Compound flood studies (e.g. Nasr et al., 2021; Wahl et al., 2015; Ward et al., 2018) that only estimate the dependence structure between pairs of stations are less affected by these data gaps and do not attempt to infill missing observations. As this study also investigates dependence across locations, constraining the analysis to common time periods without missing data would likely result in very few overlapping events. We calculate the length of the 41-year observational data with removed gaps. The data length sharply decreases from 41 years to 11.4 years for the West coast and to 3.2 years for the combined Gulf and East coasts when only gap-free records are used, see Table S3. Using such short overlapping data would be insufficient to robustly estimate the dependence structure.
>
> To address this issue, other studies (e.g. Jane et al., 2020; Quinn et al., 2019) infill data gaps or missing values using simultaneous values from nearby stations. This prepares complete time series for dependence estimation, but this approach may introduce artificial signals such as increased correlation between locations. To preserve sufficient data coverage across locations, we infill the missing values in the time series at those 41 combinations of tidal gauges and river stations. Across all locations, the averaged infilling percentage is 1.73% (i.e. equivalent to 0.71

years) for the 41 observation years between 1980 and 2020, see Table S3. For daily maximum total water levels, each of the 41 tidal gauges has missing values, with 33 gauges missing less than one year of data. Two gauges, Santa Monica and Bar Harbor, show the lowest data completeness, with 3.2 and 3.6 years of missing values respectively. For daily river discharges, 10 stations contain missing values where six stations have gaps of less than one month, two stations have missing data up to two years, and one station (Cowlitz River) is missing 7.5 years of data.

Table S3: Comparison of data lengths between original data with removed gaps and complete data with infilled gaps for the West and the combined Gulf and East coasts. Numbers in the brackets show the average percentage of infilled data across all stations.

Method	West	Gulf + East
Constrained to commonly available events (excluding missing values)	11.4 years	3.2 years
Complete time series (infilling missing values)	41 years (1.73%)	41 years (1.73%)

3. L161-186, as well as Section 2.4, provide a description of critical methodological steps of this work; however, they lack a rigorous mathematical framework. The authors should include suitable mathematical formulations, possibly using subscripts for referring to generic locations or events, when explaining the proposed approach.

We agree. We have expanded the description of our methodological steps by including the necessary mathematical formulations and a schematic exemplifying the construction of spatially joint events, see below.

L181: These bivariate extreme events at individual locations are then grouped into a large dataset for each study region. To do this, we consider a set of m locations (i.e. 13 and 28 station combinations for the West and the combined Gulf and East coasts). At location i, we use a bivariate vector $X_i = (TWL_i, Q_i)$ where TWL_i and Q_i represent time series of paired total water level and river discharge. The set of these components for each study region is then defined as $\boldsymbol{X} = \{X_i,\ i \in \{1, \dots, m\}\}$.

We further transform $\boldsymbol{X}$ onto a common marginal scale. Laplace margins are adopted in this study because they have been shown to outperform other common marginals such as Gumbel distributions in the subsequent dependence modelling framework (Keef et al., 2013). For the set $\boldsymbol{X}$, the transformation is achieved by:

$$Y_i = \begin{cases} log\{2F_i(X_i)\}, X_i < F_i^{-1}(0.5) \\ -log\{2[1 - F_i(X_i)]\}, X_i \geq F_i^{-1}(0.5) \end{cases} \quad (1)$$

where F_i is the marginal distribution of X_i. The marginal distribution F_i is semi-parametric and estimated independently per component at individual locations. For each water level or river discharge component, a generalised Pareto distribution (GPD) is fitted to detrended and de-clustered peak values peak values above a specified threshold while an empirical distribution is used for those below the threshold. We use the previously identified thresholds for this process and the underlined GPD fitting is performed through penalised likelihood estimation using a Gaussian prior. To assess the sensitivity of the transformation results to different

marginal distributions; we also test Gumbel marginals and find that the results are insensitive to this choice.

L194: From each transformed set $\boldsymbol{Y_{trans}} = \{Y_i,\ i \in \{1, \ldots, m\}\}$, we identify spatially joint events across the entire coastal region, see Figure 2 for an example of constructing one such event for a region with 7 locations. To do this, we first identify the primary variable with the largest marginal value (e.g. the water level at location 4, marked in orange) among all variables from the entire dataset, and retrieve the occurrence date and location. At this primary location, we then obtain the corresponding value for the other variable (e.g. the river discharge in the hatched orange cell) from the sampled bivariate events. For instance, if the largest extreme water level event occurs at a coastal station, we obtain the corresponding river discharge value at the paired river station from the bivariate event set developed for individual locations.

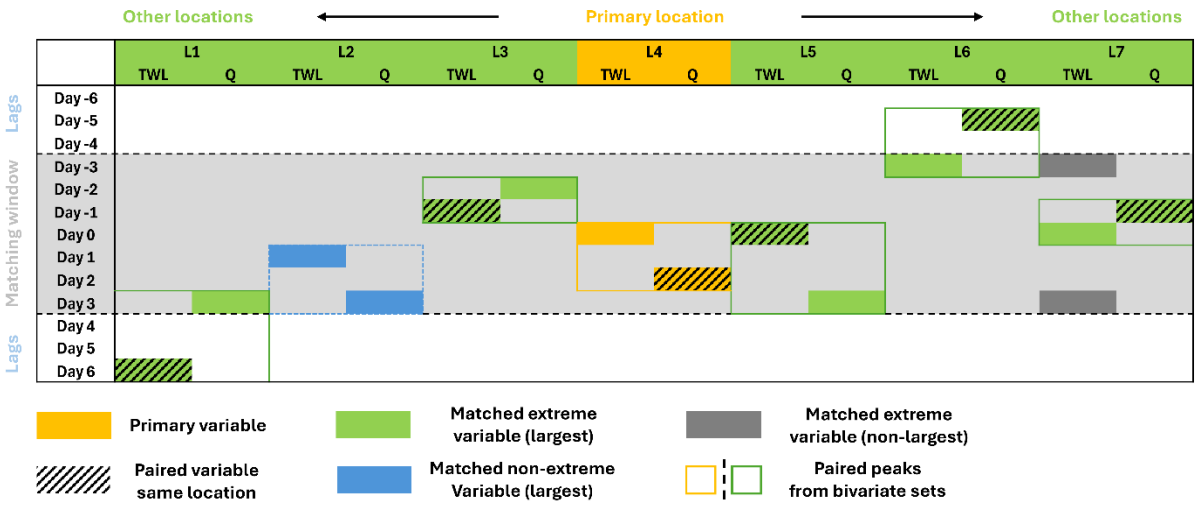

Figure 2: Schematic of the construction of one spatially joint events across the 7 locations of an exemplary coastal region. SWL refers to total water level while Q refers to river discharge. The orange cell indicates the primary variable with the largest marginal value at the primary location. For other locations, matched extreme variables are marked in green where the maximum of either TWL or Q within the matching window is selected. Dark grey cells are the available extremes within the window but they are not the largest. Hatched cells are the paired peaks to the matched variable from the bivariate events identified for individual locations. Blue cells indicate the matched non-extreme variables; they are not from paired bivariate peaks and are marked by a blue dashed box.

L213: This process may result in multiple bivariate events identified for a single location (e.g. the three extreme water level identified at location 7); in these cases, we retain the event with the largest marginal peak (e.g. the event marked by a green square at location 7). If no event is found for a particular location (e.g. the case for location 2), we instead select the maximum total water level or river discharge (e.g. the blue cells at location 2) within the 7-day window.

L222: This approach generates a separate dataset $\boldsymbol{Y}$ of spatially joint events of total water level and river discharge from the large transformed dataset $\boldsymbol{Y_{trans}}$ with time series of paired peaks for the two study regions in this study.

L244: The multivariate conditional model works by 1) estimating the univariate marginal distribution for each variable; and 2) calculating the pairwise dependence structure based on regression functions. We use the same marginal distributions $\boldsymbol{X}$ as estimated in Sect. 2.3. To estimate the dependence between total water levels and river discharges across different locations, we apply the multivariate conditional model to the transformed datasets of identified spatially joint extreme events $\boldsymbol{Y} = \{Y_i,\ i \in \{1, \dots, m\}\}$ (Sect. 2.3).

L262: The estimated dependence structure (Sect. 2.4.1) describes the conditional distribution of variables at other locations when one of the two variables (i.e. total water level and river discharge) at a given location is extreme. This information can be used to develop an event set of a large number of spatially co-occurring events, whereby for individual events at least one variable at one location is extreme.

We apply a Monte Carlo procedure to generate a 10,000-years of spatially joint events of total water levels and river discharges across different locations for each study region. We denote this 10,000-years event set as $E = \{\boldsymbol{y} \in \mathbb{R}^{\boldsymbol{d}}: \exists i \in \{1, \dots, m\}, y_i > u\}$, where u is a high threshold. We use the same thresholds as identified in Sect. 2.3. The event set E can be separated into subsets of events conditioned on a given variable following $E_i = \{\boldsymbol{y} \in \mathbb{R}^d: y_i > u \text{ and } y_i = \max(\boldsymbol{y})\}$. To quantify the number of events to be generated for each subset E_i, we use a multinomial distribution with the total event number n_s of the 10,000 years and a probability vector $P(\boldsymbol{Y} \in E_i | \boldsymbol{Y} \in E)$ for $i \in \{1, \dots, m\}$.

To construct the multinomial distribution, we first calculate the empirical distribution of annual event counts using the dataset of identified spatially joint extreme events $\boldsymbol{Y}$ (Sect. 2.3). For the 10,000 simulation years, the total event number n_s is approximated by summing up 10,000 values randomly sampled from the annual event count distribution. The next step is to estimate the probability vector $P(\boldsymbol{Y} \in E_i | \boldsymbol{Y} \in E)$ for $i \in \{1, \dots, m\}$. From the identified spatially joint extreme events $\boldsymbol{Y}$, we obtain the conditioning variable for each event, which is defined to have the largest marginal value among all variables. We then calculate the likelihood of each variable being the conditioning variable. The probability vector is then combined with n_s to calculate the event number n_i for each subset E_i. Lastly, E_i is generated by repeating the following simulation steps until n_i is satisfied.

4. While addressing the previous comment, the authors should clarify if and how they account for mixed-population effects within each water-level or flow time series. In several parts of the manuscript (e.g., L288, L355, L368, L415), the authors observe that there are different possible generating mechanisms for these extremes, such as tropical cyclones and extra tropical cyclones affecting storm surges, or snowmelt and convective events affecting riverine flooding. Could the coexistence of different parent distributions limit the applicability of the proposed methodology? The authors should discuss these aspects in detail.

Thank you for raising this good point. The proposed framework does not account for mixed-population effects and instead assumes that all extreme events originate from a single population. This simplification is primarily because addressing those effects would require stratifying historical events into distinct categories, such as TC-, ETC-driven events as well as those generated by hydrological processes like snowmelt and convective rainfall. Such event stratifications require long data time series (e.g., the 122-year observations used in Maduwantha et al., (2024)), which are not yet available for large-scale analyses. We recognise that this is one of the key limitations of this study. We have reflected

on this point and have suggested potential ways that could be used to account for mixed-population effects for future studies.

> L176: When identifying these extreme events, we follow previous studies (e.g. Couasnon et al., 2020; Ghanbari et al., 2021; Jane et al., 2020; Wahl et al., 2015; Ward et al., 2018) and assume that all events arise from a single population. This simplifying assumption therefore does not account for the mixed-population effects caused by events generated by different storms (e.g. TCs and ETCs) and hydrological processes (e.g. snowmelt and convective rainfall).
>
> L550: In regions where compound flooding can result from multiple synoptic weather patterns (e.g. TCs and ETCs) and hydrological processes (e.g. snowmelt and convective rainfall), different generation mechanisms may produce distinct dependence structures between flood drivers (Kim et al., 2023). To capture these mixed-population effects, our stochastic event generation could be improved by distinguishing events based on their generation types rather than combining all events into a single population (c.f. Maduwantha et al., 2024). Such event stratifications require long and continuous time series of flood drivers (e.g., the 122-year observations used in Maduwantha et al., (2024)), which may not be available for large-scale analyses. Future work could therefore consider using additional datasets for a long time series synthetic TCs (e.g. Bloemendaal et al., 2020) and ETCs derived from seasonal forecasting data (e.g. Benito et al., 2025), as well as hydrological data generated by stochastic weather generators (e.g. Falter et al., 2015; Ullrich et al., 2021).

5. L402: it is expected that larger storms will have a greater spatial footprint and affect a larger number of stations. The authors should therefore rephrase the sentence at L402. For example: "However, these relative occurrence rates become significantly higher with increasing thresholds, reflecting the fact that bigger storms can affect more locations"

Thank you for pointing this out. The sentence has now been rephrased into:

> L450: With increasing thresholds, these relative occurrence rates become significantly higher. This is primary because larger storms are expected to have a greater spatial footprint and may therefore affect more locations.

6. The code provided to reproduce the analyses (assets for the review process) apparently lacks several parts, including a routine for automated threshold selection (L154), and subroutines that are called within the provided scripts. Those subroutines contain functions that are used in the shared scripts; without those files, it is impossible to run the provided codes. E.g., in the R script "Cal_dependence_autmoated_thres.R", calls to the following missing scripts are included: "transFun.HT04.R", "predict.mex.conditioned.R", "u2gpd.R", "revTransform.R", "Migpd_Fit.R", "mexTransform.R".

Thank you for your careful review of the scripts and for raising this important point. We have included all the necessary functions and codes in the repository. In addition, both the dataset and code repositories are now publicly accessible to facilitate a better understanding and broader adoption of the proposed approach and developed datasets.

The dataset can be retrieved at: https://doi.org/10.5281/zenodo.15728000; and the scripts to generate this dataset and figures related to this submission can be found at: https://doi.org/10.5281/zenodo.17464793.

L591:

Data availability

The dataset containing 10,000 years of spatially joint events of extreme sea levels and river discharges along the U.S. coastlines is publicly available on Zenodo: https://doi.org/10.5281/zenodo.15728000 (Li, 2025a) under the Creative Commons Attribution 4.0 International license.

Code availability

The scripts to develop the synthetic dataset and produce the figures in this manuscript are provided at https://doi.org/10.5281/zenodo.1746479 (Li, 2025b), which are publicly accessible under the Creative Commons Attribution 4.0 International license.

Minor comments

1. L49-50: add references supporting this statement regarding the effects of TCs and ETCs on extreme and moderate, more frequent events, respectively.

Thank you. We have added references to this statement, see below:

L49: While TCs tend to cause extreme flooding, ETCs are found to be responsible for more frequent and moderate events (Booth et al., 2016; Gori & Lin, 2022).

2. Fig 2b, L293-296: How were the 5th-95th-percentile confidence intervals obtained?

Thank you. For the observed data, the 5^{th}-95^{th} confidence intervals were calculated using the rl.evmOpt.R function in the texmex R package. This function assumes a simple symmetric confidence interval based on the estimated approximate standard errors from 1,000 random observations. For the simulated data, the confidence intervals were estimated by taking the 5^{th} and 95^{th} percentiles from 100 model realisations of 1,000 years of values randomly sampled from the 10,000-year event sets.

We have added this information in Section 2.4.3 of Data and Methodology, as follows:

L291: The simulated return values are the median return levels obtained from 100 model realisations, each representing 1,000 years of randomly sampled values from the 10,000-year event sets. For each realisation, return levels are calculated empirically using the Weibull's plotting formula (Makkonen, 2006). We also estimate the 5^{th}-95^{th} confidence intervals (CIs) for both observed and simulated return levels. For the observed values, symmetric CIs are computed based on the estimated standard errors from 1,000 random samples. For the simulated return levels, the CIs are given by the 5^{th} and 95^{th} percentiles of the estimated return levels from the 100 realisations.

3. Fig 4 is unclear. Does the y-axis show the percentage of events affecting 1, 2, …, etc. distinct locations over the 10,000-simulation period? If yes, please consider using a more informative label for the y-axis.

Thank you for your good suggestion. Yes, the y-axis indeed shows the percentage of events with associated affected locations over the 10,000-year simulation period. We have changed the label into *Percentage of simulated events*.

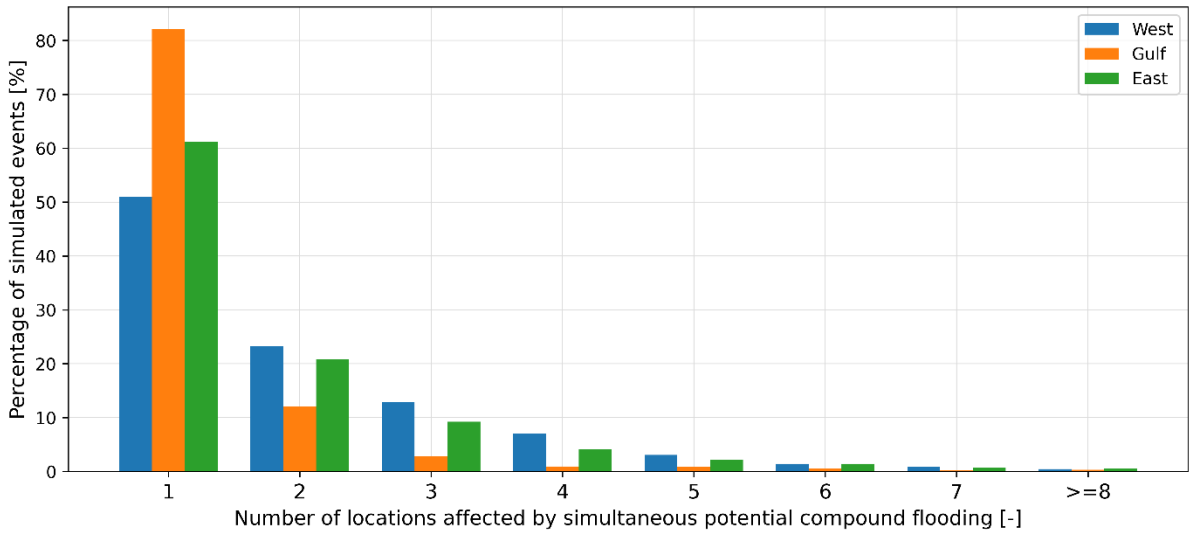

Figure 5: Event percentage diagram with the number of locations affected by simultaneous potential compound flooding for the West Coast, Gulf of Mexico, and East Coast. Potential compound flooding is defined by events where both total water level and river discharge exceed their respective 99th percentiles. Over the 10,000-year simulation period, the total number of potential compound flooding events is 24,086, 15,540, and 28,635 for the West, Gulf, and East coasts, respectively.

References

Benito, I., Eilander, D., Kelder, T., Ward, P. J., Aerts, J., & Muis, S. (2025). *Pooling Seasonal Forecast Ensembles to Estimate Storm Tide Return Periods in Extra-Tropical Regions*. https://www.authorea.com/users/903251/articles/1278266-pooling-seasonal-forecast-ensembles-to-estimate-storm-tide-return-periods-in-extra-tropical-regions

Bloemendaal, N., Haigh, I. D., de Moel, H., Muis, S., Haarsma, R. J., & Aerts, J. C. J. H. (2020). Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Scientific Data 2020 7:1*, *7*(1), 1–12. https://doi.org/10.1038/s41597-020-0381-2

Booth, J. F., Rieder, H. E., & Kushnir, Y. (2016). Comparing hurricane and extratropical storm surge for the Mid-Atlantic and Northeast Coast of the United States for 1979–2013. *Environmental Research Letters*, *11*(9), 094004. https://doi.org/10.1088/1748-9326/11/9/094004

Couasnon, A., Eilander, D., Muis, S., Veldkamp, T. I. E., Haigh, I. D., Wahl, T., Winsemius, H. C., & Ward, P. J. (2020). Measuring compound flood potential from river discharge and storm surge extremes at the global scale. *Natural Hazards and Earth System Sciences*, *20*(2), 489–504. https://doi.org/10.5194/nhess-20-489-2020

Falter, D., Schröter, K., Dung, N. V., Vorogushyn, S., Kreibich, H., Hundecha, Y., &..., & Merz, B. (2015). Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain. *Journal of Hydrology*, *524*(June), 182–193. https://doi.org/10.1016/j.jhydrol.2015.02.021

Ghanbari, M., Arabi, M., Kao, S.-C., Obeysekera, J., & Sweet, W. (2021). Climate Change and Changes in Compound Coastal-Riverine Flooding Hazard Along the U.S. Coasts. *Earth's Future*, *9*(5), e2021EF002055. https://doi.org/10.1029/2021EF002055

Gori, A., & Lin, N. (2022). Projecting Compound Flood Hazard Under Climate Change With Physical Models and Joint Probability Methods. *Earth's Future*, *10*(12), e2022EF003097. https://doi.org/10.1029/2022EF003097

Jane, R., Cadavid, L., Obeysekera, J., & Wahl, T. (2020). Multivariate statistical modelling of the drivers of compound flood events in south Florida. *Natural Hazards and Earth System Sciences*, *20*(10), 2681–2699. https://doi.org/10.5194/nhess-20-2681-2020

Keef, C., Papastathopoulos, I., & Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, *115*, 396–404. https://doi.org/10.1016/j.jmva.2012.10.012

Kim, H., Villarini, G., Jane, R., Wahl, T., Misra, S., & Michalek, A. (2023). On the generation of high-resolution probabilistic design events capturing the joint occurrence of rainfall and storm surge in coastal basins. *International Journal of Climatology*, *43*(2), 761–771. https://doi.org/10.1002/joc.7825

Maduwantha, P., Wahl, T., Santamaria-Aguilar, S., Jane, R., Booth, J. F., Kim, H., & Villarini, G. (2024). A multivariate statistical framework for mixed storm types in compound flood analysis. *Natural Hazards and Earth System Sciences*, *24*(11), 4091–4107. https://doi.org/10.5194/nhess-24-4091-2024

Makkonen, L. (2006). Plotting Positions in Extreme Value Analysis. *Journal of Applied Meteorology and Climatology*, *45*(2), 334–340. https://doi.org/10.1175/JAM2349.1

Nasr, A. A., Wahl, T., Rashid, M. M., Camus, P., & Haigh, I. D. (2021). Assessing the dependence structure between oceanographic, fluvial, and pluvial flooding drivers along the United States coastline. *Hydrology and Earth System Sciences*, *25*(12), 6203–6222. https://doi.org/10.5194/hess-25-6203-2021

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., & Heffernan, J. (2019). The Spatial Dependence of Flood Hazard and Risk in the United States. *Water Resources Research*, *55*(3), 1890–1911. https://doi.org/10.1029/2018WR024205

Ullrich, S. L., Hegnauer, M., Nguyen, D. V., Merz, B., Kwadijk, J., & Vorogushyn, S. (2021). Comparative evaluation of two types of stochastic weather generators for synthetic precipitation in the Rhine basin. *Journal of Hydrology*, *601*, 126544. https://doi.org/10.1016/j.jhydrol.2021.126544

Valle-Levinson, A., Dutton, A., & Martin, J. B. (2017). Spatial and temporal variability of sea level rise hot spots over the eastern United States. *Geophysical Research Letters*, *44*(15), 7876–7882. https://doi.org/10.1002/2017GL073926

Wahl, T., Jain, S., Bender, J., Meyers, S. D., & Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, *5*(12), 1093–1097. https://doi.org/10.1038/nclimate2736

Ward, P. J., Couasnon, A., Eilander, D., Haigh, I. D., Hendry, A., Muis, S., Veldkamp, T. I. E., Winsemius, H. C., & Wahl, T. (2018a). Dependence between high sea-level and high river discharge

increases flood hazard in global deltas and estuaries. *Environmental Research Letters*, *13*(8), 084012. https://doi.org/10.1088/1748-9326/aad400

Ward, P. J., Couasnon, A., Eilander, D., Haigh, I. D., Hendry, A., Muis, S., Veldkamp, T. I. E., Winsemius, H. C., & Wahl, T. (2018b). Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries. *Environmental Research Letters*, *13*(8), 084012. https://doi.org/10.1088/1748-9326/aad400