Reviewer 1

The authors would like to thank the reviewer for their constructive comments that have strengthened this manuscript.

The reviewer's comments have been numbered using the format QX.Y where Q is the reviewer query, X is the reviewer comment number and Y represents a subdivision of the comment if there are multiple items requiring attention. The authors' response to each comment is given below each question/comment using the format RX.Y.

Q1.1 Their definition of the compound event is wrong. The authors have sampled the largest OWL and the largest precipitation event within a given year without considering a short/limited time window. Typically for compound floods, it is assumed that the time between each driver is from 0 to a few days, preferably within a week to account for synoptic circulation. Since the 'time' of elapsing of one event to another is sufficiently large here – it merely resulted in a bivariate/multivariate event series rather than a compound pair. Likewise, pairing within a wet season month does not contribute towards compound event construction.

R1.1 The authors refer to Senevirante et al. (2012), Leonard et al. (2014), and Zscheischler et al. (2018) for the definition of a compound event. Coinciding sampling would represent compound events while, the maximum sampling strategy would not. In order to accommodate both event sampling types, the authors have adopted the reviewer's suggestion and updated the manuscript to reflect "multivariate" rather than compound events.

Q1.2 Although they quote a certain publication from FEMA pointing 'observations within a wet season' as worst possible pairs - In fact, FEMA (2015, 2018) do not provide any unified guidelines for analyzing compound floods. In FEMA (2020), a recommendation for combined occurrence for coastal and riverine flooding is provided, however, no sampling methods were discussed. This is in reference to Lines 110 to 115 in the manuscript.

R1.2 The authors have substantially updated the text to provide additional context on the sampling and have removed any reference to "worst-case" or worst possible pairs. The new paragraph reads:

Multivariate flood probabilities are determined with combinations of sampling methods: Annual Maximum (AM), Annual Coinciding (AC), Wet Season Monthly Maximum (WMM), and Wet Season Monthly Coinciding (WMC). AM sampling pairs the single largest precipitation and OWL observations within a given year (without regard to co-occurrence), where AC sampling pairs the single largest precipitation observation within a given year to the largest OWL observation within its 24-hour accumulation period. Each sampling method samples from a unique probability space and therefore will provide varying perspectives for a return period. A summary of each sites' associated gauges, observation windows, and number of pairs is provided in Table 2. Southern California's wet season is defined between October to March and provides a majority of the total annual rainfall (Cayan and Roads, 1984; Conil and Hall, 2006). It is likely for extreme multivariate events to occur during this period. Maximum sampling pairs the single largest precipitation and OWL observations within each year or wet season month. A multivariate event created with the largest observed precipitation and OWL within a year or wet season month can result in an event with severe flooding potential. Although strictly speaking maximum parings (annual or wet season) do not technically represent an observed multivariate event, they would represent a severe event and are consistent with the blended approach recommended by FEMA (2016c). Coinciding sampling pairs the single largest precipitation observation within each year or wet season month to the largest OWL observation within its 24-hour accumulation period, providing more realistic pairs compared to maximum sampling.

Q2.1 The authors have used AIC and ML methods as a goodness-of-fit estimate to select copula. However, the AIC and ML method does not provide any goodness-of-fit statistics of the population fit and the power of the test - it compares the empirical vs theoretical fit of the data. For the selection of copulas, a double parametric bootstrap method is to be adopted, other than the AIC-based criteria (Genest et al., 2009; Rémillard, 2017).

R2.1 The authors have performed the analysis using a Cramér-von Mises tests (e.g., Genest et al., 2009, Couasnon et al, 2018, Sadegh et al., 2018, Ward et al., 2018). The manuscript has been updated and revised accordingly.

Q3.1 In compound event perspective, it is of interest to analyze when the design level exceeds both variables simultaneously, and not either of the variables in isolation. Therefore, the 'OR' criteria is suitable to analyze multivariate probability and not the compound extreme per se. The papers they have followed, also emphasize multivariate events and estimation of associated hazard and therefore, do not explicitly consider compound events in particular.

R3.1 The authors agree and have revised the manuscript to reflect “multivariate event”. Please also see response **R1.1**.

Definitions of a compound event

- “(1) two or more extreme events occurring simultaneously or successively; (2) combinations of extreme events with underlying conditions that amplify the impact of the events; or (3) combinations of events that are not themselves extremes but lead to an extreme event or impact when combined. The contributing events can be of similar (clustered multiple events) or different type(s)” (IPCC 2012, Senevirante et al. 2012)
- The combination of processes (climate drivers and hazards) leading to a significant impact is referred to as a ‘compound event’. (Zscheischler et al., 2018)
- The combination of variables or events that lead to an extreme impact is referred to as a compound event. (Leonard et al., 2014)

References

- Couasnon, A., Sebastian, A., and Morales-Nápoles, O.: A copula-based Bayesian network for modeling compound flood hazard from riverine and coastal interactions at the catchment scale: An application to the Houston Ship Channel, Texas, Water, 10, 1190, 2018.
- FEMA: Guidance for Flood Risk Analysis and Mapping: Coastal Flood Frequency and Extreme Value Analysis, https://www.fema.gov/sites/default/files/2020-02/Coastal_Flood_Frequency_and_Extreme_Value_Analysis_Guidance_Nov_2016.pdf, 2016.
- Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, Insurance: Mathematics and economics, 44, 199–213, 2009.

- Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford-Smith, M.: A compound event framework for understanding extreme impacts, Wiley Interdisciplinary Reviews: Climate Change, 5(1), 113-128, 2014.
- Sadegh, M., Moftakhari, H., Gupta, H. V., Ragno, E., Mazdiyasni, O., Sanders, B., Matthew, R., and AghaKouchak, A.: Multihazard scenarios for analysis of compound extreme events, Geophysical Research Letters, 45, 5470–5480, 2018.
- 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.
- Zscheischler, J., Westra, S., Van Den Hurk, B.J., Seneviratne, S.I., Ward, P.J., Pitman, A., AghaKouchak, A., Bresch, D.N., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events, Nature Climate Change, 8(6), 469-477, 2018.

Reviewer 2

The authors would like to thank the reviewer for their constructive comments that have strengthened this manuscript.

The reviewer's comments have been numbered using the format QX.Y where Q is the reviewer query, X is the reviewer comment number and Y represents a subdivision of the comment if there are multiple items requiring attention. The authors' response to each comment is given below each question/comment using the format RX.Y. Comments are address in the order they appear from the reviewer.

Q1.1 As pointed out by the authors, "Serinaldi (2015) suggests inter-comparing univariate, multivariate, and conditional probabilities and return periods is misleading as each probability type describes its associated event." (line 337-338). Yet, after reading the manuscript it gives the impression that this is exactly what the authors did by listing the results obtained from all possible hazard scenarios. In the methodology, the authors mention the type of question each hazard scenario might be answering, but this is not reflected in the abstract. Instead, only a broad mention of "compound events" is mentioned.

R1.1 The abstract and manuscript have been substantially updated in accordance with the reviewer's suggestion. The abstract now specifically refers to hazard scenarios (Line 6). Lines 327-335 clarify when and how each scenario may be applied. Similarly, additional context was added at Lines 335-337 to emphasize the purpose of exploring all possible hazard scenarios within this study. Please refer to the redline text for the associated updates.

Q2.1 The authors also frequently refer to the FEMA (2016) methodology which seems to imply that they want to compare their hazard scenarios with FEMA but this is not done at any point in the manuscript. FEMA (2016) provides a methodology to derive the 100 (500)- year return period of the water level while this study does not model the obtained water levels from the event pairs obtained. Furthermore, the FEMA flood maps should represent, in theory, the flood depths levels happening once every 100 (or 500) years on average. This probability cannot be compared with probabilities from the hazard scenarios and conditional probabilities as each of them only contains a subset of all potential flood events.

R2.1 The currently recommended FEMA methodology represents a simple and efficient methodology relying on annual maximum or peaks over threshold sampling to consider flood hazards. Nonlinear interactions may amplify flood risk in areas where multi-source flooding occurs (e.g., Moftakhari et al., 2019). Currently, FEMA does not provide guidance on methods estimating flood risks associated with the co-occurrence of precipitation and high water levels. This study explores the influence of sampling and copula selection on precipitation and water level pairs. Although it is outside the scope of work presented here, future work will hydrodynamically model precipitation-water level pairs and quantitatively compare flood extent to univariate, blended models similar to the FEMA methodology. The manuscript has been edited to avoid any suggestion that these hazard scenarios were hydrodynamically modeled and directly compared to FEMA flood maps.

Q3.1 It is unclear what is meant by "worst-case" in the manuscript which makes the abstract and conclusions very difficult to understand. Based on which results do the authors conclude that AM does not result in worst case scenarios? Is this for each hazard scenario?

R3.1 All references to "worst-case" has been removed in the manuscript. Annual maximum sampling pairs the largest observations within a year creating a data set of multivariate events with relatively large OWL and precipitation. Generally, the annual maximum sampling provides the largest OWL values. However, the 100-year return period for OR and SK scenarios exhibited larger OWL values using water year monthly coinciding sampling (WMC; Figure 9b, c) as discussed in the paragraph starting on Line 269. The paragraph stating on Line 269 in the manuscript has been expanded and revised to provide additional context.

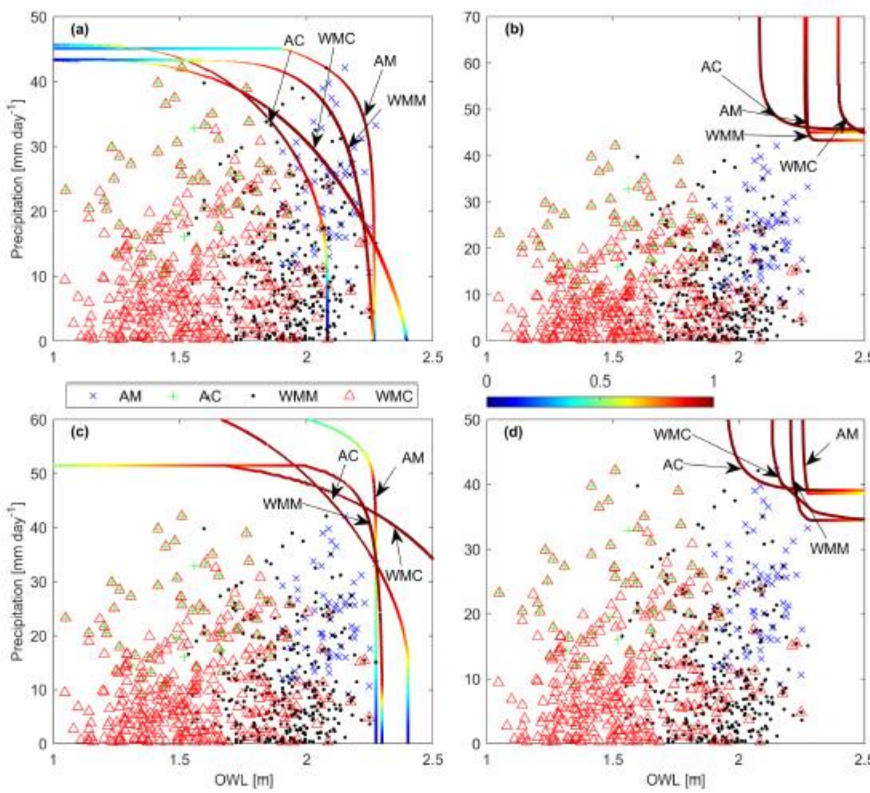

Figure 9. San Diego (a) AND, (b) OR, (c) SK, and (d) K hazard scenarios for annual maximum (AM, cross), annual coinciding (AC, plus), wet season monthly maximum (WMM, dot), wet season monthly coinciding (WMC, triangle) data and 100-year isolines using the BB1 copula. Sampling labels point to the mostly likely value on their respective isolines.

This behavior was also observed in the marginal and conditional cases (paragraph starting Line 257) and led to the conclusion that annual maximum sampling may not result in the largest potential flood events at long return periods. Additional text was added to provide further clarification to Line 276 which reads "Given wet season monthly coinciding sampling results in larger OWL values for the marginal, conditional, OR, and Kendall scenarios, this suggests maximum type sampling may not accurately reflect OWL at extreme return periods".

Q4.1 To my understanding, this study does not assess flood risk but derives the pairs of different joint/conditional return periods, except for the structural scenario where a mention on the flood extent is given. As such, I find the title a bit misleading in its current form. Also, the implications of the work could be better highlighted in the abstract. I find parts of the discussion interesting (lines 339-348) where the authors mention the value of testing different hazard scenarios to capture compound interactions.

R4.1 The authors have amended the text to consider events rather than flood risk. The title has been updated to reflect the reviewer's suggestion and is now "Characterizing multivariate coastal flooding events in a semi-arid region: The implications of copula choice, sampling, and infrastructure". The abstract has been updated

Q5.1 Line 9-10: "Although annual maximum sampling is commonly recommended for characterizing compound events, …". The current emphasis in the abstract about the annual maximum sampling is not straightforward to me. This could also be because I am not sure what the authors mean by "worst case" event pairs. But I would actually tend to disagree with this statement. There is a wide variety of sampling methods commonly used to characterize compound floods that include both peaks-over-threshold and annual maxima, with no clear scientific consensus for one or the other method. Compound flood studies that have looked at the impact from these events, such as Santos et al. (2020) found AM methods to produce a more balanced datasets. However, without looking at the impact from your events, one cannot make a strong conclusion about this.

R5.1 The authors have substantially updated the abstract and manuscript to remove any references to "worst-case" and provide a more balanced approach to potential sampling methods. Please refer to the redline text for changes. The authors agree there is not a scientific consensus for one sampling method or the other and provide multiple studies utilizing both methods in the paragraph starting on Line 50. "Recommended" has been changed to "used" to prevent suggestion of any particular sampling method. Santos et al., (2021) recommends monthly maximum for more balanced datasets. In this case, the monthly maximum paring exhibited inverse relationships owing to the dry season where essentially no precipitation occurs. As an alternative, results are presented for the wet season monthly maximum.

Q6.1 I find the comparison of different sampling methods particularly confusing (section 4.3 and 4.4). As mentioned by reviewer 2, these methods sample different sets. Therefore, it is expected to lead to different values for the evaluation of the '100-year' probability. When presenting the results, this should be clearly discussed, otherwise this seems to imply that '100-year' event from the marginal, conditional and joint probabilities are the same, which is not true. Maybe this could already be clearly stated in the Methods and reiterated in the Results. In the end, an important message from your study is specifically that these hazard scenarios do not represent the same probability space and thus for the same probability level, they lead to different values. This means that extreme caution should be used for users/modelers who want to use a given hazard scenario to derive the "100-year" flood event.

R6.1 The authors agree that the various sampling methods represent different probability spaces and should result in different return period values. The variation among sampling spaces has yet to be explored within the literature. Text was added on Line 109 and Line 257 to provide context and clarify this point. The added text reads: "Each sampling method samples from a unique probability space and therefore will provide varying perspectives for a return period" and "It should be noted that each sampling method represents a unique probability space and accordingly results in alternative realizations of a given return period. ".

Q7.1 The authors did not show formal goodness-of-fit test (but only informal ones like the BIC and AIC). The lowest BIC or AIC value could still denote a poor fit from the model. It is therefore usually recommended to perform the Cramer-von Mises blanket test.

R7.1 The authors have updated the analysis using the suggested Cramér-von Mises tests (e.g., Genest et al., 2009, Couasnon et al, 2018, Sadegh et al., 2018, Ward et al., 2018), and the manuscript updated accordingly. Although the specific copula selections shifted, the overall findings are similar. Please refer to the updated Methods, Results, and Discussion sections within the redline manuscript to see the respective changes.

Q8.1 Line 1: Your study assumes stationary conditions so I am not sure the sentence on sea level rise is needed.

R8.1 The sentence was removed per the reviewer's suggestion.

Q9.1 Line 38: "and then adopting the more severe flooding result". From the reference cited (FEMA 2016c), it is shown that FEMA does not uses the most severe flood result but a combination of both fluvial and coastal flood depth probabilities to estimate the 100-year water level in transitional areas (see Figure 4.1 in the FEMA Guidance for Flood Risk Analysis and Mapping: Coastal Flood Frequency and Extreme Value Analysis). So I would rephrase this to better highlight the limitations from their methods.

R9.1 The authors thank the reviewer for pointing out this error. Lines 35 to 37 have been amended and now read, "For example, FEMA recommends characterizing multivariate events by developing univariate water level and discharge statistics and then adopting a smoothed, blended result for transitional areas (FEMA, 2011, 2016c)."

Q10.1 Line 50-51: Maybe rephrase to mention that initial studies were mainly focusing on hazard scenarios. I would then advise to change the title of Table 1 to a non-exhaustive list of studies that used hazard scenarios (if this is what the authors want to emphasize). Otherwise, the body of literature that used copulas to study compound flooding is much broader that what is stated there. Also, there is a lot of recent research that moved beyond single hazard scenarios in order to better capture all potential flood events using copulas.

R10.1 The reviewer's suggestion was adopted**.** Line 58 in the updated manuscript now reads as "Initial studies were primarily focused on select hazard scenarios…" and Table 1's caption will now start as, "A non-exhaustive list of multivariate studies…", per the reviewer's comments.

Q11.1 Line 67: "humid climatic conditions": This sounds very broad

R11.1 The authors have added specific information to constrain the term "humid climatic conditions". Line 67 now includes a reference to Koppen-Geiger system, Beck et al. (2018).

Q12.1 Line 75 : "most likely". At this point, this term has not been introduced in the manuscript and without context, this may lead to the wrong understanding of this expression.

R12.1 The authors thank the reviewer for pointing out this subtle but important point. The text has been updated to remove "most likely" and replace with "resulting".

Q13.1 Line 92: "December 19, 2023" Typo?

R13.1 The authors thank the reviewer for catching this typo and changed “2023” to “2013”.

Q14.1 Line 93-94: I am not sure of the added value of this sentence. Can you make a clearer link with your study design?

R14.1 From an oceanographic prospective, the term tide specifies water levels changes only from astronomical influences. Observed water level represents tide and any storm surge (i.e., inverse barometric effects and wind/wave setup) or climatic influences (such as El Nino). A sentence was added at Line 92 to clarify the importance of the distinction between observed water level and tide. The additional text reads: “This is a key distinction since compound event dependencies may change depending on water level selection.”

Q15.1 Line 103: “by the event time”. Do you mean duration? Or time step?

R15.1 “Event time” has been updated to “event duration” in accordance with the reviewer’s suggestion.

Q16.1 Line 114-116: “.. since the co-occurrence of precipitation and water levels follows the FEMA guidance for considering a “worst case scenario" approach .” I do not understand this. Can you elaborate?

R16.1 The language “worst-case” has been removed and new text added to Line 115. The added text reads “Although strictly speaking maximum parings (annual or wet season) do not technically represent an observed multivariate event, they would represent a severe event and are consistent with the blended approach recommended by FEMA (2016)”.

Q17.1 The authors use a lot of acronyms. I can understand why but the references to the acronym is not always mention in the Figures or titles. So the reader sometimes has to scroll to the text to find those acronyms again. For example, I couldn’t find what M refers to (I think Marginal distribution?). I would carefully review all Figures and Tables to make sure any acronym is mentioned to be able to interpret the figure or table independently of the text.

R17.1 All tables and figures have updated captions that provide utilized acronyms.

Q18.1 Table 3: The word “marginal” or “univariate” when describing the probabilities/return period is missing in the title.

R18.1 The probabilities in Table 3 are for all scenarios (marginal, conditional, bivariate) utilizing Wet Season Monthly Maximum or Wet Season Monthly Coinciding samplings. Probabilities of exceedance (or their inverse, the return period) describe a probability threshold or front (in the bivariate case) dependent on sampling (Eq. 15 in the manuscript) and the available data (length, observations, spread, etc.). The return period is not unique to the marginal case, and therefore not specified as such. Table 4 provides the fitted marginal distributions and Figure 3 displays the marginal CDFs.

Q19.1 Line 194: “easily”. I would remove this term and rename this section “Univariate return periods”. In multivariate space, I think to interpret a return period is not easy at all!

R19.1 The term “easily” was removed and the sentence now reads “Return periods (T) provide a metric describing the severity of an event…”.

Q20.1 Line 320: “save the Cubic”: typo?

R20.1 This line was removed after significant revision required using Cramer-von Mises test.

Q21.1 Uncertainties are not quantified in your study which limits the interpretation of the implications. I am not asking to add this but at least mention this point somewhere in the discussion.

R21.1 The line "The impact of sampling and distribution choice on uni- and multivariate return period values are presented, however uncertainties deserve further characterization." was added on Line 369.

Q22.1 Similarly, it would be interesting to mention the limitation of this method in locations with multiple pluvial or coastal flood seasons.

R22.1 The reviewer's suggestion has been adopted. Line 370 to 372 was added to address the limitations associated with multiple flood seasons. "This work focused solely on exploring conditional and joint probabilities of OWL and precipitation in a tidally and wave dominated semi-arid region and would not be applicable to regions experiencing multiple flooding seasons (e.g., Couasnon et al., 2022)".

References

- Couasnon, A., Sebastian, A., and Morales-Nápoles, O.: A copula-based Bayesian network for modeling compound flood hazard from riverine and coastal interactions at the catchment scale: An application to the Houston Ship Channel, Texas, Water, 10, 1190, 2018.
- Couasnon, A., Scussolini, P., Tran, T., Eilander, D., Muis, S., Wang, H., Keesom, J., Dullaart, J., Xuan, Y., Nguyen, H., et al.: A flood risk framework capturing the seasonality of and dependence between rainfall and sea levels–an application to Ho Chi Minh City, Vietnam, Water Resources Research, p. e2021WR030002, 2022.
- FEMA: Guidance for Flood Risk Analysis and Mapping: Coastal Flood Frequency and Extreme Value Analysis, https://www.fema.gov/sites/default/files/2020-02/Coastal_Flood_Frequency_and_Extreme_Value_Analysis_Guidance_Nov_2016.pdf, 2016.
- Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, Insurance: Mathematics and economics, 44, 199–213, 2009.
- Moftakhari, H., Schubert, J. E., AghaKouchak, A., Matthew, R. A., and Sanders, B. F.: Linking statistical and hydrodynamic modeling for compound flood hazard assessment in tidal channels and estuaries, Advances in Water Resources, 128, 28–38, 2019.
- Sadegh, M., Moftakhari, H., Gupta, H. V., Ragno, E., Mazdiyasni, O., Sanders, B., Matthew, R., and AghaKouchak, A.: Multihazard scenarios for analysis of compound extreme events, Geophysical Research Letters, 45, 5470–5480, 2018.
- Santos, V. M., Wahl, T., Jane, R., Misra, S. K., and White, K. D.: Assessing compound flooding potential with multivariate statistical models in a complex estuarine system under data constraints, Journal of Flood Risk Management, 14, e12 749, 2021.
- 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.