

## ***Interactive comment on “Multivariate statistical modelling of the drivers of compound flood events in South Florida” by Robert Jane et al.***

**Robert Jane et al.**

r.jane@ucf.edu

Received and published: 10 July 2020

### **Response to Anonymous Referee 1**

Authors have proposed a framework for assessing the compounding effects of flooding drivers in coastal areas of South Florida (e.g. rainfall, coastal water level (WL) and groundwater level). They first assess the significance of dependence between these variables, and then propose bivariate and trivariate approaches to generate compound design scenarios. They finally compare the generated design scenarios with the current design approaches (with the assumption of full dependence between variables). The idea is very interesting and manuscript is very well written. This work would be a significant and novel

C1

contribution to the community and deserves publishing in NHESS, however after a major revision. My main concerns are: i) continuity of probability density function along the hazard isolines must be checked, ii) generated bivariate and trivariate hazard scenarios are significantly inconsistent at the margin (rainfall-WL plane) which needs a careful consideration. Details are provided below:

The authors would like to thank the reviewer for providing thoughtful comments. Please find our responses to each comment below including how we plan (following the NHESS review process) to adjust the manuscript.

#### **Major comments:**

- When the quantile-isoline is created by overlapping two separate isolines, could you please explain how you ensure that the joint PDF estimates are consistent along the envelope? I mean, theoretically taking derivative of CDF function, we are not necessarily dealing with a continuous function, and there is a chance that derivatives (that give us relative likelihood estimates and shading along the curve in Figure 5) diverge around the break point, right? Please, elaborate here. I see that you explained the process as “In this work, the relative probabilities are estimated nonparametrically via a Kernel Density Estimate (KDE), using the ks R package (Duong, 2007). Initially KDE was applied to the observations, however, particularly for larger return periods the design event proved highly sensitive to a small number of observations. Hence, design events were determined by applying KDE to a large sample = 10,000 from the two fitted copulas, with sample proportions consistent with the empirical distributions, and transformed back to original scales.” But not yet clear to me how you check the continuity of PDF along the quantile-isoline. This probably helps better understand why “For small return periods (20 years), design event rainfall remained <1mm, thus they may be considered “surge only” events.” Looking at Figure 5a, we see some orange spectrum around the break point, while most likely scenario falls on the mar-

C2

gin! You have also come up with the conclusion that “At sites S20 and S28, although the bivariate design events for return periods 1- and 100-years were “surge events”, non-negligible probability density was located along part of the isoline comprising compound events” To me it raises a reddish flag that continuity is omitted.

The derivation of the isolines and the estimation of the contours are carried out independently. For the latter, as discussed in the manuscript and quoted above, a kernel density estimate is used to estimate the joint PDF of the observations (approximated by large samples from the two fitted copulas). Once a given isoline is derived the probability density associated with pairs of rainfall and O-sWL falling on the isoline are extracted (and scaled to lie within  $[0, 1]$ , hence relative probabilities). Therefore, no differentiation of the CDF is required and continuity is preserved. To clarify that no differentiation of the CDF is required as the probabilities are obtained directly from the KDE the following sentence has been added to L387:

*“The probability density given by the KDE at points along the isoline are extracted and the probabilities scaled to lie within  $[0, 1]$ , hence yielding relative probabilities.”*

- In page 18, where you analyze the timing before the O-sWL in the bivariate design event derived from the two-sided sampling approach reaches the corresponding value obtained from SFWMD scenarios (Figure 7), there is a significant underlying assumption that needs an explicit explanation, which is non-stationarity of the correlation structure. Indeed, the Copula parameters are assumed to remain unchanged over time.

The authors agree this is an important assumption which warrants an explicit explanation in the text. The following passage has been added to the manuscript.

*“Nonstationarity in the dependence between rainfall and O-sWL can occur as a consequence of a range of anthropogenic and climatic induced stressors. In this study, the dependence is assumed stationary i.e., that the copula parameters remain unchanged over time. The overlapping records at the three sites are of insufficient length to ro-*

C3

*bustly test the stationarity assumption. However, Wahl et al. (2015) did not detect any significant change in Kendall's  $\tau$  between rainfall and surge at either Key West or Mayport, the two closest sites to Miami-Dade County, indicating stationarity may be a reasonable assumption. Nevertheless, due to regional and local effects, such as multi-decadal variation in the storm surge climate, the possibility of statistically significant trends in the dependence cannot be ruled out at the case study sites.”*

**There is significant difference between compound scenarios using bivariate and trivariate approaches. In Figures 5a and 5b, the design scenario is picked so close to the margin (with rainfall approx. 0) and a conclusion is made “In the bivariate analysis, at site S22, low return period ( $< 20$  year) design events constituted “surge only” events”. While in Figure 8 upper right panel, RP= 10 yr comes with 200 mm of rainfall. Also, in Figure 5c the pair associated with RP=50 yr is (Rainfall = 185, WL = 1.11), while in Figure 8 (upper right panel) the pair associated with RP = 50 yr is (Rainfall = 340, WL = 1.778) that is significantly larger than the one proposed under bivariate analysis. I understand that sampling approach is different between cases, but distribution functions must be compatible at the margins (i.e. at Rainfall-WL plane) otherwise yields in a great confusion. Especially when you state “The output of the bivariate and particularly trivariate applications can also act as boundary conditions for coupled hydrologic/hydraulic models for assessing flood risk and designing flood defence structures”. Also, bivariate approach is based on extreme two-sided sampling, which is a more realistic approach compared with the one under trivariate analysis. So, does this significant different between outcomes suggest inappropriateness of employed trivariate scheme with the aim of “accounting for actual dependencies”?**

Thank you for this very useful comment. As you correctly state the sampling approaches differ between the bivariate and trivariate analyses. The purpose of the bivariate analysis is to demonstrate the conservative nature of the full dependence assumption between rainfall and O-sWL used in the original design of the structures.

C4

To ensure only the impact of accounting for the dependence is being assessed, the two design events (triangles and diamonds) are derived using identical marginals. The fact that the marginal distributions are the same is evidenced in Figure 5 by the aligning of the design events derived under the assumption of full dependence (triangles) with the isolines at the margins. However, the disadvantage of using identical marginals is that the design events derived under the assumption of full dependence in Figure 5 may not be equivalent to those derived using the original design approach as that approach did not use the two-sided conditional sampling. To make this clear reference to “*the current design approach*” in this section have been changed to “*the assumption of full dependence*”.

In the trivariate analysis section the design events under the assumption of full dependence are shown in the top row of Figure 8, thus the design events in the upper right panel are likely to be more reflective of those adopted by the SFWMD using their current level of service assessments. In relation to the comparisons made in the comment, the events in the upper right panel of Figure 8 would be more comparable to the diamonds in Figure 5 rather than the triangles, however, they were derived using different marginal distributions and so any comparison must be made with care. The design events derived under the assumption of full dependence (first row of Figure 8) and realizations from the trivariate models (remaining rows of Figure 8) are obtained using identical marginal distributions, hence the distribution functions are compatible at the margins and the effect of accounting for the dependence can be assessed. Nevertheless, this was not explicitly discussed in the original manuscript, therefore the following sentence has been added on L489 to address this.

*“Overall the sparsity of simulation data near the design events (with return periods greater than 1-year) obtained under the assumption of full dependence demonstrates the importance of accounting for the dependencies between the drivers when assessing the compound flood hazard.”*

A comparison between the bivariate and trivariate results using the new methods

C5

(conditional sampling in the two-dimensional case and copulas/HT04 in the three-dimensional case) would involve comparing the bivariate design events i.e., diamonds in Figure 5, with the realizations of rainfall-O-sWL from the trivariate models i.e., red dots in the 2nd-4th row of the 3rd column of Figure 8 (as opposed to the first row of Figure 8). As correctly stated in the comment the bivariate approach yields “surge only” events for lower return periods at site S-22. Realizations from the trivariate models with low O-sWL (say 1mNGVD) and zero rainfall are common particularly for the Gaussian and vine copulas in agreement with the bivariate results. For the HT04 method which is arguably the most flexible and thus robust of the trivariate methods, as the O-sWL increases events with zero rainfall become increasingly rare. The HT04 model thus implies higher return period events are more likely to become compound in nature as return period increases, again, in common with the bivariate results. The latter trend is less pronounced for the Gaussian and vine copulas nevertheless the disparity between the bivariate and trivariate results is perhaps not as great as implied in the comment. As correctly stated in the comment sampling approaches are different between the (bivariate and trivariate) cases and the distribution functions must be compatible at the margins if the ability of the bivariate and trivariate approaches to capture the dependence are to be compared. Furthermore, the two cases are addressing different issues consequently we refrain from comparing the two cases in the manuscript.

To summarize the response so far, both the trivariate and bivariate cases the marginal distributions used to derive the design events with/without dependence are identical in order isolate the effect of modeling the dependence.

**Minor comments:**

**L180-185: Sampling specifications not provided.**

Thank you for the comment. The first sentence of Section 3.2 has been expanded to give more details about the sampling specifications and now reads as follows:

*“Here, a two-sided sampling approach similar to that in Wahl et al. (2015), which*

C6

*involves deriving two conditional samples where each variable is conditioned on in turn, is implemented to identify bivariate extreme events.”*

In addition, a reference to the results section where more details on the choice of thresholds can be found has also been added to the end of the paragraph starting on L182:

*“For more details on the choice of thresholds see Section 4.2.”*

**L223: Define the variables used in this equation.**

The variables are now defined in the text.

**L496: have not defined “mNGVD”**

Thank you for pointing this out, a definition of mNGVD has been added.

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2020-82>, 2020.