

Interactive comment on “Evaluating the efficacy of bivariate extreme modelling approaches for multi-hazard scenarios” by Aloïs Tilloy et al.

Anonymous Referee #2

Received and published: 23 March 2020

This study evaluates the efficacy of various bivariate models for a variety of correlation structures in a large set of synthetic data. While I am not a fan of synthetic data, the rigorous and comprehensive evaluation of various models justify the publication of this paper in NHESS. However, there are some issues that remain to be resolved before final publication:

1. There is a large focus on multi-hazard analysis in this paper. However, I cannot see the sector that is impacted by the presented case studies. For a multi-hazard coastal flood analysis, for example, it is intuitive how higher water levels and larger river flows lead to a larger risk of flooding. However, who is impacted by rainfall and wind? What is the final impact that is worsen by concurrence and sub-sequence of multiple hazards? Even more confusing is the wildfire case study. Wildfire is the impact

C1

of high-temperature, hence I am not clear whether it is justified to consider the wildfire and temperature as two drivers of one impact. What is going to be that impact? How is it worsen by the combination of these two hazards? The manuscript should justify the presented multi-hazard analyses.

2. Wildfire burned area is related to temperature not the number of fires. A fire of size 0.1 ha can occur all year around, specially for human started fires (see Balch, J. K., et al. (2017). Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences*, 114(11), 2946-2951.). But more importantly, what is the impact? If it is the wildfire, is it justified to have fire as the impact and as the hazard? Is high-temperature necessarily a hazard?

3. Abstract needs some improvement. I struggled to understand how the case studies are related to synthetic data.

4. An important missing element in the evaluation of models is p-value. It would be interesting to see what the p-values are and determine whether the models fail/pass to represent the data! This is actually very important. The metrics used in the paper are subjective - although valuable - and a more widely accepted metric could help the general audience relate the study to other modeling practices.

5. There are many typos in the text. I highlighted some of them in the attached manuscript, but there are more.

6. There are some specific comments that are provided in the attached manuscript.

Hope the authors find the comments useful to improve their paper

Please also note the supplement to this comment:

<https://www.nat-hazards-earth-syst-sci-discuss.net/nhess-2020-28/nhess-2020-28-RC2-supplement.pdf>

2020-28, 2020.

C3