Interactive comment on “Towards a compound event-oriented climate model evaluation: A decomposition of the underlying biases in multivariate fire and heat stress hazards” by Roberto Villalobos-Herrera et al.

Anonymous Referee #1

Received and published: 7 December 2020

Relying on copula theory, the authors propose a multivariate bias assessment to separate biases in hazard from univariate drivers from their dependence. This is a relevant topic to understand compound events and how climate models can represent and simulate them or not. This framework is applied to two hazard indicators related to fire (Chandler Burning Index, CBI) and heat stress (wet-bulb globe temperature, WBGT) hazards. Overall, the paper is interesting and well constructed, with appropriate statistical experiments developed in Bevacqua et al., 2019. I enjoyed reading this submission. It clearly deserves to be published. Nevertheless, I have some comments that the au-
thors may want to include in their manuscript. Those are mostly for clarifications and general consideration of the methodology.

Specific comments.

- L. 33-34: “well-designed physically based multivariate bias adjustment should be considered for hazards and impacts that depend on multiple drivers”. I fully agree on this sentence that recurrently come back in the “bias correction” literature. However, it is not clear what this (“physically based adjustment)” means in practice, as no physically based bias adjustment is suggested (and we can argue that this is also the case in the literature). Do the authors have some in mind?

- Here, only inter-variable dependence (between T and RH) is considered. Can this framework be applied or extended to deal with temporal dependence (e.g., dependence between a variable and the same variable with a given lag) or spatial dependence? Or even both?

- In the same idea: here, each indicator is made only of 2 variables (temperature and relative humidity). Does the proposed framework work in 3 or 4 variables? I.e., in a higher dimensional context?

- In a context with more than two variables, I guess that the choice of the “non-parametric framework” (i.e., empirical distributions) is not appropriate anymore. This needs to be more discussed (although already briefly mentioned in the “discussion” section).

- Fig 1: the figure seems to show (visually) that the biases visible in panel c) mostly come from strong biases in the marginal (i.e., panels a and d) and not really from the dependence structure in panel b) that seems equivalent for ERAI and BNU. Is that correct? If so, I am not convinced that this model is the best example, as it would have been more informative/illustrative to show results for a model where both (marginal and dependence) contribute to the biases in the bivariate distribution.
L.204-205 and 226-228: “if the model sample value of $\tau$ lies within the confidence interval calculated for its corresponding ERA-Interim sample, the model sample is judged to not significantly differ from ERA-Interim in terms of the rank correlation between T and RH.” and “Like our evaluation of Kendall’s $\tau$, if the model index lies outside the confidence interval we consider the model has a significantly different representation of extreme values of CBI and WBGT from ERA-Interim.” This is a good approach that is accepted. However, one can wonder why not testing the other way around? (i.e., testing if ERAI lies in the interval from the model). Would this give equivalent results? Please, expand.

Minor/technical comments.

- L. 85: all analysEs
- L. 93-95: “we carry out the analysis on the de-correlated time series, which are obtained from the original through subsampling every N=9 days, this is the minimum lag required to remove the autocorrelation in T and RH time series data (at 95% confidence level)”. Could the authors elaborate on this? E.g., how is “N=9 days” determined?
- L.115: As mentioned, “e” depends on T and RH but this link should be reminded in a few more details.
- L.200 and after: “$z_{\alpha}$” is not defined.