Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-383-AC1, 2021

Natural Hazards and Earth System Sciences
Discussions

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.

Roberto Villalobos-Herrera et al.

r.villalobos-herrera2@newcastle.ac.uk

Received and published: 24 February 2021

Reply to Reviewer #1 comments

We thank the Reviewer for the time they spent to review the manuscript. The helpful and constructive comments have contributed to improve the paper. The text below contains our responses to each comment.

General comments: 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 authors may want to include in their manuscript. Those are mostly for clarifications and general consideration of the methodology.

We thank the Reviewer for the input and the positive feedback on the paper.

Specific comments:

C1: 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?

With this phrase we meant to refer to the design of the overall procedure, also including the climate model selection, which is taken up in the discussion ("Climate model output should be a reliable input for the bias adjustment methods, e.g., models should provide a plausible representation of large-scale atmospheric circulation (Maraun et al., 2016; 2017).")

To make this clearer we will modify the abstract, using the word "procedure": "well-designed physically-based multivariate bias adjustment procedure"

And we will further modify the discussion, to explicitly refer to the above: "These findings exemplify the need for multivariate bias adjustment methods, which can adjust climate model biases in the dependencies between multiple drivers of hazards (Francois et al., 2020; Vrac, 2018). Furthermore, relying on climate models that plausibly represent large-scale atmospheric circulation (Maraun et al., 2016; 2017) would improve our confidence in the simulation of multivariate hazards."

C2-4: - 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 "nonparametric framework" (i.e., empirical distributions) is not appropriate anymore. This needs to be more discussed (although already briefly mentioned in the "discussion" section).

While the paper focussed on disentangling biases in individual physical drivers of compound events, it is in principle possible to extend the framework to higher dimensions. Cases with more than two driving variables require the use of multivariate copulas (e.g., vine copulas to decompose the dependencies between variables or other methods to build the multivariate model). The same applies for incorporating temporal dependencies. Overall, considering more dimensions adds complexity and would require larger sample sizes but does not limit the method to the bivariate case. To help clarify this we will add the following to the discussion at the end of L.384:

"... For example, in the case of three variables X1, X2, and X3, we would have to investigate the behaviour of marginals and then the dependence between X1 and X2 (with the 2-Copula C12), X2 and X3 (C23), and X1 and X3 (C13), and then joint behaviour of the three variables with the 3-Copula (C123). Similar considerations apply for the consideration of temporal dependencies. The analysis can be done using both a parametric or non-parametric approach. For instance, in Vezzoli et al. (2017), a non-parametric approach has been used to analyse the behaviour of the three variables precipitation, temperature and runoff."

The bias decomposition framework for each individual variable only relies on the empirical distributions of model and reference observations and thus does not change for cases with more than two variables.

C5: - 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.

Thank you very much for the suggestion. We agree that an illustrative example of the bias contribution from both the marginals and the copula would be more appropriate for this figure. In the revised manuscript, we will modify Figure 1 by replacing the model BNU-ESM with the model IPSL-CM5A-LR, which presents bias contributions from both the marginals and the copula (see proposed figure at the end and in the attached pdf). In addition, and in following with the comments by Reviewer 2, we will simplify the caption to:

Figure 1: Copula-based conceptual framework employed in this study to evaluate biases in CBI and WBGT indices. The framework is illustrated for a representative location in Brazil (Amazon, 5°S and 56.5°W; indicated via X markers in the next figures). Panel (c) shows the bivariate distribution of T and RH based on ERA-Interim (grey) and IPSL-CM5A-LR data (black) during 1979-2005. Isolines indicate equal levels of CBI (orange) and WBGT (green). The decomposition of biases from the marginals (a, d) and the copula (b) are illustrated as the discrepancies between the black (IPSL-CM5A-LR model) and grey features (ERA-Interim).

C6: - L.204-205 and 226-228: "if the model sample value of τ 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 τ , 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.

Thank you for this interesting question. Testing the other way around would be computationally much more expensive as confidence intervals would need to be calculated for every model, and in addition it would produce results which would be more difficult to interpret. We thus consider our approach preferable and all models are tested against the same reference.

MINOR COMMENTS:

- L. 85: all analysEs

Thank you, this has been corrected.

- 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?

We agree that greater detail would be helpful here and will add the following details to the manuscript: "... we carry out the analysis on the de-correlated time series, which are obtained from the original through subsampling every N=9 days, where N is the lag required to remove the autocorrelation in T and RH time series data everywhere (at 95% confidence level). The value of N was determined as follows: for all grid points and years in ERA-Interim and the CMIP5 models, the autocorrelation function was calculated; then, the minimum lag for which the autocorrelation was non-significant at the 95% confidence level was determined. Finally, the maximum of all the minimum

lags was selected, resulting in N=9 days. The time series for all models and locations are sampled with the frequency of N. This is done N-times using different start epochs, where the first sampled time series starts with time epoch one, the second sampled time series with time epoch two and so on up to nine. The de-correlated time series of T and RH will henceforth be simply referred to as samples in the following sections."

- L.115: As mentioned, "e" depends on T and RH but this link should be reminded in a few more details.

We are not sure how to interpret this comment. Here we simply state what is evident from the equation used to compute e, namely that it depends on air temperature and relative humidity.

- L.200 and after: "z_{\alpha}" is not defined.

This is the quantile of the standard normal distribution; the following will be added to the paper: "where σˆ2 is an estimator of var(τ) and z_(α/2) is the quantile of the standard normal distribution for α/2 (Hollander et al., 2014)."

Please also note the supplement to this comment:
https://nhess.copernicus.org/preprints/nhess-2020-383/nhess-2020-383-AC1-supplement.pdf

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-383, 2020.

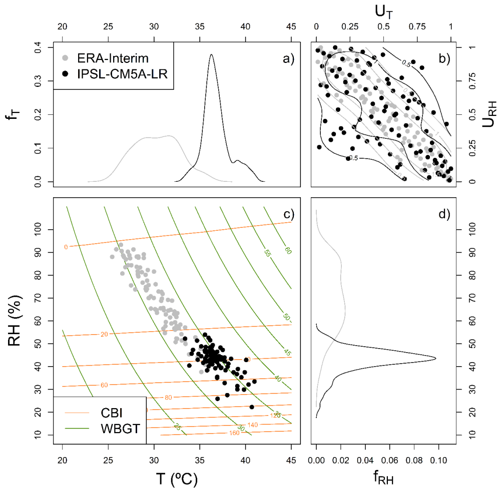

Figure 1: Copula-based conceptual framework employed in this study to evaluate biases in CBI and WBGT indices. The framework is illustrated for a representative location in Brazil (Amazon, 5ºS and 56.5ºW; indicated via X markers in the next figures). Panel (c) shows the bivariate distribution of T and RH based on ERA-Interim (grey) and IPSL-CM5A-LR data (black) during 1979-2005. Isolines indicate equal levels of CBI (orange) and WBGT (green). The decomposition of biases from the marginals (a, d) and the copula (b) are illustrated as the discrepancies between the black (IPSL-CM5A-LR model) and grey features (ERA-Interim).

Fig. 1. Modified figure 1, see Reviewer 1 comment 5