Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-383-AC2, 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 #2 comments

Many thanks to the Reviewer for their comments and the time and effort that required to prepare them. We believe your input will contribute to improving our paper. You may find our responses to your comments below.

General comments: The study by Villalobos-Herrera et al. deals with an important topic and offers an approach to identify and quantify the source of biases in multivari-

ate impact-based indicators derived from climate model simulations. The methods are correct and well explained, although there are few clarifications needed (see the specific comments). Results are effectively presented and useful to start reflecting on this complex topic. Concerning the manuscript, I found the discussion session a repetition of concepts already discussed and again mentioned in the conclusions. I suggest to make it shorter and more focused on the key point, that to me is the difficulty in having multivariate bias adjustment methods and the complexity of some impact indicators based on several different variables.

Thank you for your positive feedback. We have re-written a portion of our discussion, in particular L370-L384 to highlight the importance of multivariate adjustment methods and hazard indicator complexity. This will be implemented in the revised manuscript:

Our results underline the importance of attributing the sources behind biases in multivariate hazard indicators beyond simple univariate assessments. Biases in multivariate hazard indicators can be rather complex, as exemplified by our two example indicators which show very different bias structures despite being constructed by the same variables. Climate models that tend to simulate too high T also tend to simulate too low RH and vice versa (Fischer and Knutti, 2013), this behaviour would be expected to cause compensating biases in WBGT and enhance biases in CBI. In fact, biases in WBGT are smaller than the bias contributions from T and RH, demonstrating the presence of compensating biases for WBGT. A negative inter-model correlation between the contributions of T and RH to WBGT biases reduces the biases in WBGT in the CMIP5 average. While we have found a positive inter-model correlation between the bias driven by T and RH, no enhancement of the CBI bias occurs because CBI is mainly controlled by RH (see isolines in Figure 1c), which consequently also controls the bias of the index.

Specific comments: C1: Figure 1 caption is 'chaotic'. I suggest to re-write.

We have modified Figure 1 according to comments by Reviewer 1 (see figure at the

bottom and in the attached supplement). The caption has been modified to read:

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).

C2: Figure 2 is not very informative. I would either improve or remove.

We believe Figure 2 facilitates reader understanding of the data processing procedure, however we agree that it may be superfluous in the main text and will move it to the Appendix.

C3: Figure 7 I would modify Panel b to let readers better appreciate the identified behaviour.

Figure 7b has equal axes for both sources of CBI bias (as does Figure 9b). We opted for this choice such to highlight the contrast between the small spread and magnitude of the T bias contribution relative to that of RH. We believe that having axes with different ranges would make this less evident. To help clarify this aspect we will add the following sentence to the figure caption:

Equal axes are used in (b) to highlight the differences in spread between both bias components.

C4: L87: this implicitly means you assume models are able to correctly reproduce the seasonality. It may be worth to discuss it.

Thank you for this suggestion, we agree and will add the phrase to our text:

"Following Zscheischler et al. (2019), we restrict our analysis to the hottest calendar month of the year, which is selected based on the climatology of ERA-Interim data at each grid point. This choice was made because arguably heat stress and fire hazards tend to be more frequent during the warmest period of the year, and it avoids dealing with seasonality, however we note that this assumes that CMIP5 models correctly reproduce the seasonality observed in ERA-Interim."

C5: L96 more details should be added on this estimated lag.

Thank you, we agree, the explanation will be expanded as shown below:

"... 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."

C6: L100-115 add a brief explanation on all these absolute numbers, just to let readers better understand.

Both the employed indices are widely used in the scientific literature. The parameters were defined within the original sources. Given the comment of the referee, we will add a sentence at the end of the section: "... More details on the definitions of the CBI and WBGT are available at McCutchan and Main ,1989; and ACSM ,1984"

C7: L157 following your notation Uerai is the transformed random variable (unif distributed) from Terai.

This is correct, we will clarify this in the text: “From the variable Terai we calculated the uniformly distributed transformed random variable UT,erai =FT, erai(Terai).”

C8: L191-192 Since many tests exist to compare distributions, I do not understand this sentence on K-S and A-D. I would delete it.

We agree and will remove this sentence.

C9: L213 Here, on the contrary, I would add an explanation. Why CvM test and not the A-D you use for marginals? As they both belong to the same test family.

Thank you for the interesting question. Our goal is here to test whether two bivariate empirical copulas are equal. Hence, we use the test by Remillard and Scaillet (2009), which was specifically developed for this task. In general, we note that if we used a multivariate version of the A-D test and the null hypothesis (of equality in distribution) was rejected, then we would have no way of knowing whether the difference in the bivariate distributions was due to a difference in the marginal distributions or a difference in the dependence structure (i.e. the copula) or both.

To clarify why we use this test will add some text to the paper (in square brackets below):

“Note that different copulas may give rise to the same value of τ, therefore we cannot conclude that a model that faithfully reproduces the ERA-Interim values of τ is accurately representing the full dependence structure between T and RH. Therefore, we account for [differences] in the dependence structure by also carrying out hypothesis tests which are based on the full copula function. We perform the non-parametric test of copula equality based on the Cramer-von-Mises test statistic proposed by Remillard and Scaillet (2009),” used in Vezzoli et al. (2017) for testing the capability of a climate-hydrology model to reproduce the dependence between temperature, precipi-

tation and discharge for the Po river basin in Italy, and recently employed by Zscheischler and Fischer (2020) for evaluating the ability of climate models to represent the dependence between temperature and precipitation in Germany. The copula equality test has a null hypothesis of H0: Cerai = Cmod where Cerai and Cmod are the copulas of T and RH represented in ERA-Interim and a given model respectively, with the alternative hypothesis being that these copulas differ. [Unlike the AD test, which can evaluate CMIP5 model performance in reproducing a single marginal distribution, the copula equality test was specifically developed to test whether two empirical copulas are equal and thus evaluates the capacity of models to reproduce the full dependency structure between T and RH.] We used the TwoCop function of the TwoCop R-package (v1.0, Remillard and Plante, 2012) to run the test.”

C10: L319 According to the Figure, it seems that the dependence contributes much less than the others.

We agree, as we discuss later in the paragraph. We open the paragraph at line 319 with an introductory sentence where we do not give details on the contributions of the copula and the two marginals. We feel this is warranted as the contribution of the dependence varies in space. This is discussed, together with the comment of the referee, at L330:

“In addition, we observe a tendency towards a lower bias, on average, driven by the copula component (global area weighted average of absolute bias equal to 0.85°C); note that, however, some relevant positive bias contributions exist over eastern Brazil and central Africa where the copula test shows higher frequencies of rejection (Figure 5c), and negative contributions over northern Russia, Central United States, and eastern Europe (Figure 8f). “

Please also note the supplement to this comment:
https://nhess.copernicus.org/preprints/nhess-2020-383/nhess-2020-383-AC2-supplement.pdf

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

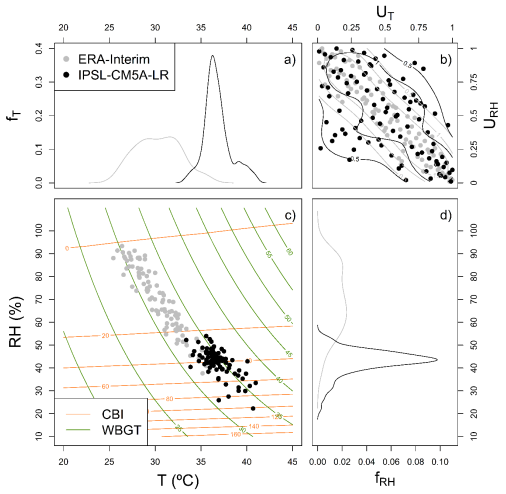

Figure 1: Copula-based conceptual framework employed in this study to evaluate biases in CBI and WBGT indices. The framework is illustrated for a representative location in Brazil (Amazon, 5ºS and 56.5ºW; indicated via X markers in the next figures). Panel (c) shows the bivariate distribution of T and 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