We thank the Stefan Hagemann for his helpful comments. Our replies below are highlighted in italic.

The authors present an interesting study on how univariate bias corrections of climate model output affect impact indicators that depend on more than one climate variable. They chose two hazard indicators related to heat stress and fire risk to demonstrate to demonstrate the effect of separate univariate bias corrections in comparison with a multivariate method. The latter corrects the dependence structure between the variables in addition to the respective variable distributions. The paper is written well so that I have only a few minor comments.

- The title seems to be too general. The focus of the paper is on impact indicators that depend on more than one climate variable, and not on impact modelling in general. I suggest revising the title.

We agree and will change the title to "The effect of univariate bias adjustment on multivariate hazard estimates".

- I suggest citing (in the introduction and discussion of results) Räty et al. (2018) who actually found that in many cases a multivariate bias correction is not necessary (from the hydrological perspective). They stated that "the additional benefit of using bivariate bias correction methods is not obvious, as univariate methods have a comparable performance. "

Räty, O.; Räisänen, J.; Bosshard, T.; Donnelly, C. Intercomparison of Univariate and Joint Bias Correction Methods in Changing Climate From a Hydrological Perspective. Climate 2018, 6, 33.

Thank you for this suggestion. Indeed, we already mention that in the hydrological context, the discussed issues are likely not very important because often precipitation is the dominant driver (p. 11 l. 19). However, even though many hydrological impacts might not be affected by incorrect multivariable dependencies, this might look different for impacts that strongly depend on multiple variables. Based on individual examples where multivariate bias adjustment did not lead to improvements, we cannot draw the general conclusion that multivariate bias adjustment is not necessary in any case. This is discussed in on p. 11 l. 4 onwards. We will include the paper by Räty et al in this discussion.

Räty et al. also point out that differences between impact simulations based on climate input data bias-adjusted with univariate versus multivariate methods are most often smaller in cross-validation than in validation metrics. This also does not imply that multivariate bias adjustment is generally unnecessary, as cross-validation can results can be strongly misleading (Maraun & Widmann, 2018). We will extend our discussion section towards this point and refer to Räty et al. in that context.

- The analysis of results (Sect. 3) takes into account relative changes in the bias (reduction by at least 50%, increase). This means that also grid points are included where the bias is small/negligible for impact purposes. Here, a low reduction in bias or even a small increase in the bias would not matter for modelling the impacts. Is there a way of setting a bias threshold that defines the 'acceptable' bias, and then consider only the noteworthy changes on biases above this threshold? This means to include only points in the analysis where the bias before or after correction is above this threshold. I think that such a discrimination is helpful to judge how problematic the application of univariate bias correction is for those biases that matter. One results of the study is that univariate bias correction cannot effectively reduce biases in multivariate hazard estimates when (iii) univariate biases are small. However, if the resulting biases in the hazard indicator are small, this will not matter for the respective cases.

We agree that in many cases the absolute bias might be small. In general it is often difficult to choose the magnitude of an acceptable bias and this will be highly context- and variable dependent. Nevertheless, we agree that this is an important point to discuss. Hence, we compute the fraction of pixels for which the bias in WBGT is larger than 1 either before or after bias adjustment. This is the case for 50-90% of the pixels, depending on the model and the metric. We then recompute figure 3 based on this subset (figure shown below). The resulting fraction of locations where bias adjustment does not achieve the two chosen benchmarks is about half of the original numbers when all locations are included. We will include this figure in the revision and discuss its implications. However, because these numbers strongly depend on the size of the accepted bias, we will keep the other figures as is.

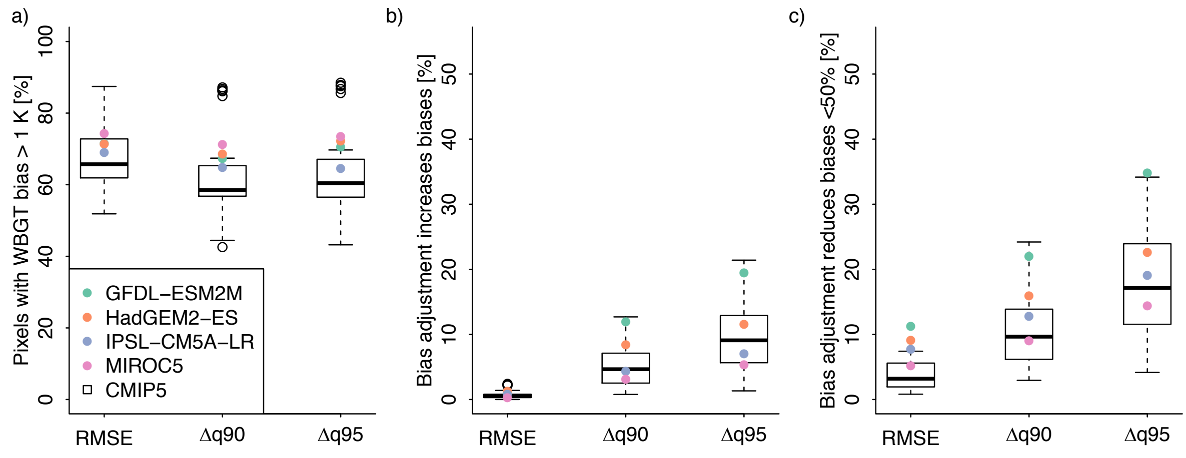

Figure: a) Fraction of pixels for which biases in WBGT are larger than 1 K before or after the application of bias adjustment. b-c) As in Figure 3 only for WBGT, based on the subset of pixels identified by a).

I suggest accepting the paper for publication after minor revisions are conducted.

Minor remarks
In the following suggestions for editorial corrections are marked in Italic.

Fig. 3
I suggest adding one line that indicate the type of each column for which WBGT and CBI are considered, i.e. RMSE, \Delta q90 and \Delta q95.

Fig. 4 and 7
I suggest using a discrete colour bar to improve the respective figures.

Fig 5.
It is difficult to identify regions in panel a). I suggest using another colour to indicate the regions, e.g. red.

p.10 – line 21
… period, *as longer* time …

p.12 – line 23
We thank Alex …

p.17 – line 20
… explain *a large* fraction …

Thank you. We will include all suggestions in the revised version.