The paper compares different bias correction approaches for correcting two heat-stress indices. Although there have been assessments of BC approaches in previous literature, including univariate and multivariate ones, the authors do offer a new perspective of comparing the direct and indirect implementations, which is often confusing for impact studies and thus worth investigating. In this regard, I believe this paper can provide useful information for the community, especially for those processing data on heatwaves or other similar compound indices. However, I have several concerns that should be addressed before the publication:

**Major comments**

- Section 2.3: The authors selected four univariate BC methods but only one multivariate method (i.e., MBCn) in this paper. Since several different MBC methods have been developed in recent years (e.g., R2D2 (Vrac 2018), MRec (Bárdossy and Pegram)), the authors may need to explain why they select MBCn here and what its characteristics are, either in the Introduction or the Method section.

Also, although you have included a detailed description of MBCn in the Supplementary Information, I suggest including general information for describing how MBC works (maybe one or two sentences) in the main text for those unfamiliar with MBC.
Figure 3: As the bias shown in the calibration and validation periods is different, the authors may consider applying the same experiments with two periods switched to see if the same systematic bias retain and how it affects the bias-corrected result. Since and apply, and whether the bias correction model changes significantly. Especially since the authors do not present future projections, using a reverse-periods experiment can increase the robustness of the result.

I am not sure how the authors could solve the problems with non-stationarity with the results of this study, which is indeed a problem of all bias correction. I suggest a discussion with a reverse-period experiment (Comment 2) to emphasize the problem in non-stationary bias, while the authors rephrase the argument with a “softer” tone.

Minor Comments:

P3, Line 88: Instead of “WBGT”, the equation (3) used in this paper should refer to “simplified WBGT”. The authors should specify this.

P7, Line 167-170: As you find almost no difference between the results of EQM and QDM in this study due to the use of only historical data, how about keeping just one of these two methods? I feel it redundant to present both here.

P8, Line 188: I think that this statement is not fully supported by the calibration period, but it’s true for the validation period. Therefore, it’s better to change the location of this sentence in the paragraph.
Reference
