

Compared with the preprint, the revised manuscript has incorporated several changes, and the overall quality of the manuscript has improved to some extent. The authors' responses to the reviewers were relatively conscientious. However, some issues raised by the editor and previous reviewers have still not been fully justified. Therefore, I think substantial revisions are still needed before this manuscript can be considered for publication in HESS.

1. The bias-correction section is still not fully convincing. The authors did not follow the editor's suggestion to provide preliminary validation. In addition, it is not common to use the mean of three reanalysis datasets as the observational reference, and they also did not quantify between-member variability as suggested by editor. Although the authors added a new paragraph introducing quantile delta mapping (QDM), they did not provide a sufficiently comprehensive discussion of how this method has been used in previous studies or whether it is appropriate for their specific case. There are many important papers in this field that were not cited.
2. The new explanations added by the authors for the Wilcoxon rank-sum test (WRST) and the climate model selection are still not convincing, and there are also some factual mistakes. The authors state that only a few models provide daily tasmax output and that the four selected GCMs offer relatively high spatial resolution. However, daily tasmax output is available for many more GCMs than suggested. In addition, the spatial resolutions reported in Table S1 do not appear to fully match the native atmospheric resolutions of the listed models, and there are other models with higher spatial resolution.
3. The editor suggested that "Adding quantitative metrics, regional analysis and deeper interpretation in the results and discussion would enhance your manuscript.", but the authors did not address this point very well. Although some additional discussion has been added, most of it remains rather superficial and mainly attempts to convey a simple message: CHWEP is more severe than a single event. This point has already been discussed many times and is, to some extent, already well recognized in the current literature (see: <https://www.nature.com/articles/s41612-024-00579-4>). Therefore, I do not think this paper currently provides enough new or sufficiently useful scientific insight to the field.