

## Comments on the manuscript: egusphere-2025-4289-ATC2.pdf

*Global scalation of more frequent and intense compound heatwave- extreme precipitation events by Jin, H. et al.*

### General comments

I thank the authors for submitting a revised version of the manuscript. While the revision shows some effort to address the reviewers' and editor's comments, several important issues raised in the previous review have not yet been fully resolved.

The initial reviews highlighted the following main concerns:

- The absence of a quantitative comparison between historical model simulations and reanalysis data over the same historical period.
- Insufficient clarity in the notation used in Equations (4) to (11).
- Sections of the results discussion remain rather superficial and do not provide the level of physical interpretation necessary to better understand the reported analyses.

In addition, the revisions concerning the bias-correction procedure, the criteria for climate model selection, and the explanation of the Wilcoxon Rank Sum Test remain limited and would benefit from further clarification and justification to meet the standards expected for publication in *HESS*.

### Specific comments

**Line 100:** The authors state that the four selected GCMs provide relatively high accuracy and spatial resolution. However, the revised manuscript still lacks a clear and systematic comparison between historical GCM simulations and reanalysis data for the same reference period. Without this comparison, the assessment of GCM performance and accuracy remains incomplete.

**Line 155:** The notation used in Equations (4) to (11) is still difficult to follow. In particular, the definition and role of the operator *mean(.)* should be explicitly clarified to improve transparency and reproducibility.

**Line 210:** The manuscript introduces a "random probability" in Equation (13). It appears that the authors aim to estimate the probability that at least one rainfall event occurs within a 7-day window following a heatwave event. However, the rationale for selecting this specific window length is not provided. Furthermore, the term "random probability" may be misleading and could be reconsidered or more clearly defined.

Reviewer 2 also noted the absence of reported  $p$ -values associated with the Wilcoxon Rank Sum Test. Additionally, the bar plots used to illustrate the proportion of grid cells with positive or negative significant/non-significant WRST results are difficult to interpret, as the color scheme is inconsistent with that used in the corresponding spatial maps.

Considering these remaining concerns, I believe the manuscript would require **major revisions** before it can be considered for publication. Addressing the points above would substantially improve the clarity, robustness, and interpretability of the study.