
Response to the RC1

Dear Reviewer,

We sincerely thank the reviewer for your care, support, and valuable assistance with our research. Your insightful comments and suggestions have been of tremendous help in continuously improving and enhancing both the content and overall quality of this study. Upon receiving your review comments, we promptly revised the manuscript in accordance with your recommendations, striving to fully meet all requested revisions. The revised portions have been highlighted in red in the manuscript for your convenient review. Our point-by-point responses to your comments are provided below.

Thank you again for your time and kind assistance.

Best regards,
Haoyu Jin

The authors present an analysis of compound heatwave extreme precipitation (CHWEP) events using statistics computed an ensemble of 3 reanalysis data (in the past) and 4 CMIP models (for projection). They first show the differences on several statistics between CHWEP and single heat waves (SH). Then, they show how these statistics vary in the future, under several SSPs.

Overall, I found that this analysis suffers from several methodological issues that are detailed below. The paper is not very well written, with important quantities not properly defined, equations I could make sense of, etc. In my opinion, this work cannot be accepted for publication in HESS, unless completely reworked.

Main comments:

1. Currently, the authors compute the statistics in the past on reanalysis and in the future on climate models. It is well known that climate models are highly biased if not corrected with bias correction methods (see eg François et al., 2020). Is it the case here?

If the answer is 'no', then I don't see how it is possible to draw any definitive conclusions when comparing the future (using CMIP models) and the past (with reanalysis) as done in Sections 4.2 and 4.3 (Figures 6 - 10 and S3 - S4). If the answer is 'yes', the authors should nevertheless first check that the WRT does not reject the equality of the distribution when the statistics are computed on the models during the past periods. This is a preliminary study that is absolutely necessary. Otherwise, one does not know for sure whether the differences observed are due to climate change or to differences between models and reanalysis.

Response: The authors applied quantile delta mapping (QDM) to bias-correct precipitation and temperature data from CMIP6 models under different future scenarios. QDM is an advanced bias correction method that preserves the climate change signal in future projections. It improves upon

traditional quantile mapping (QM) by not directly mapping future simulations onto the historical observational distribution. Instead, it transfers the modeled distributional changes between the historical and future periods, i.e., the "delta" or incremental shift, onto the observational distribution, thereby avoiding the suppression of climate change trends inherent in CMIP model projections. The Wilcoxon rank-sum test (WRST) is a non-parametric hypothesis test used to assess whether two independent samples originate from the same population (or populations with identical medians). It does not assume normality of the data and is therefore widely employed when the assumptions of parametric tests (such as the t-test) are violated. In this study, we use WRST to statistically evaluate the median differences in precipitation and temperature characteristics between compound heatwave-extreme precipitation (CHWEP) events and single extreme events (extreme precipitation and heatwaves), thereby quantitatively distinguishing CHWEP events from individual extremes. We thank the reviewer for pointing out this issue, which helps the authors further improve and refine this study.

2. Equations (1) - (8) are unclear. For example, in Eq. (1), there is a sum indexing over i, but I don't understand what the index is exactly (and there is no i in CHWEP). Also I don't understand what we are summing exactly. Eq. (3) is even less clear. What is actually computed? The mean (over all events) of the sum of the temperature (withing each events)? In this case, one should see numbers above 100? I am really lost here.

The same kind of criticism could be made for each equation. Note also that these statistics are computed on a 45 year period, across several members of an ensemble. I guess there is some sort of averaging over the years that should be made apparent.

Equation (10) is also very unclear). Why is P_{rand} a coincidence probability?

Response: The authors have revised Sections 3.2 and 3.4 to provide a clearer description of the computation of attributes for CHWEP and single extreme events, as well as the observed and random coincidence probabilities for CHWEP events. Figure 1 has also been redrawn to more clearly illustrate the extracted attribute metrics. Specifically, we first identify extreme precipitation and heatwave events using precipitation and temperature thresholds, respectively. CHWEP events are then identified by applying a 7-day temporal window to select closely successive heatwaves and extreme precipitation events. Subsequently, Equations (1)-(8) are used to calculate the attributes of CHWEP, single extreme precipitation, and single heatwave events. These attributes include:

- The number of occurrences of CHWEP and heatwave events,
- The average temperature intensity of CHWEP and single heatwave events,
- The average precipitation intensity of CHWEP and single extreme precipitation events,
- The average duration of heatwaves within CHWEP and single heatwave events.

We thank the reviewer for pointing out this issue, which has helped us further refine and improve the quality of this study.

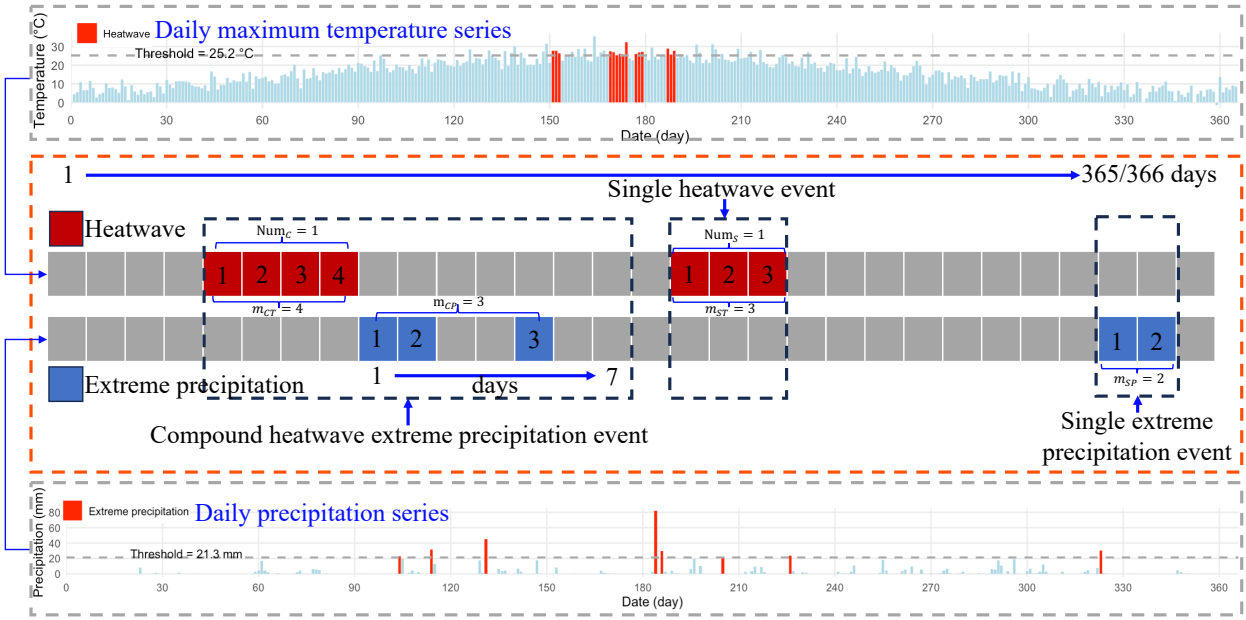

Figure 1. Schematic illustration of CHWEP and single extreme precipitation and heatwave events extracted from daily precipitation and daily maximum temperature time series over a year.

3. The authors do not explain how the ensembles are taken into account in their study. Are the above statistics computed on each member of the ensemble averaged out? Why is the number of members is different for the climate models than for the reanalysis? Does it pose a problem?

Response: In this study, to mitigate potential biases inherent in any single dataset, we adopted a multi-dataset approach by computing the mean across multiple datasets to reduce uncertainties associated with individual data sources. For the historical period (1980-2024), we used three reanalysis datasets, which are generally considered highly reliable and thus serve as benchmark datasets. For future (2056-2100) daily precipitation and daily maximum temperature data, we selected four CMIP6 global climate models that provide both variables. We first computed the ensemble mean of these four models and then applied the quantile delta mapping (QDM) method to bias-correct the model outputs. The CHWEP and single extreme events in the future period were subsequently identified using this bias-corrected, ensemble-mean data. It is worth noting that relatively few CMIP6 models provide daily maximum temperature (namely tasmax) outputs at the daily time scale, therefore, we carefully selected four models that simultaneously deliver both future daily precipitation and daily maximum temperature. We thank the reviewer for raising this point, which has helped us further refine the methodology and improve the clarity and accuracy of this manuscript.

4. The vocabulary is not consistent throughout the paper. For example, according to line 127, F_C and F_S denote frequencies (hence the F letter), but Figure 2(a) and 2(b) show counts. Note also that according to caption, Figures 2c and 2d show heatwave intensity (between 0 and 40 °C). Why "intensity", and not "temperature"? Note also that according to Eqs. (3) and (4), IT_C and IT_S are means over sums, which I expect to be above 100°C?

Response: The authors have standardized the labels in both the text and figures. Figure 2a and b show the average annual number of CHWEP and single heatwave events, respectively, from 1980 to 2024, i.e., the frequency of these events per year. Similarly, Figure 2c and d represent the average intensity of CHWEP and single heatwave events per year over the same period (1980-2024), defined as the average daily temperature during heatwave events. Using the term "intensity" aligns better with conventional practice in this field of research. We thank the reviewer for raising this point, which has helped the authors improve the accuracy of the manuscript's presentation.

5. The test used in this study is the Wilcoxon rank-test (WRT), which is consistent under specific assumptions on the two distributions (say X for a statistic computed in the past, and Y for the same statistics computed in the future). The WRT is consistent if the alternative is that Y is stochastically larger than X, i.e. $P(Y* > X*) \geq P(X* > Y*)$, where $X*$ and $Y*$ are random values from X and Y, respectively. The test on the median (line 139) is consistent with the additional assumption of that alternative is restricted to a shift in location, i.e. $F_Y(s) = \delta + F_X(s)$, with $\delta >0$.

In any case, it is misleading to state that "the WRST is emplyed to assess differences in extreme events characteristics" (line 135-136). The authors should reformulate and be more specific (and narrow).

Response: The authors have appropriately revised lines 135-136 in Section 3.3 to clarify that the Wilcoxon rank-sum test (WRST) is employed in this study to assess differences in the medians of extreme event characteristics. The WRST offers several advantages: it does not require the two samples to follow (or approximate) a normal distribution, and it imposes minimal requirements on sample size. In contrast, the t-test, used to evaluate whether the means of two groups differ significantly, requires either paired data (for the paired t-test) or independent samples that are approximately normally distributed, especially in small-sample settings, which limits its applicability. We thank the reviewer for highlighting this issue, as it has helped the authors further refine the study and improve the accuracy of the manuscript's presentation.

6. Section is almost impossible to follow; the authors compare quantities that have not been properly defined, such as standard deviations SD (line 242), "ratio of observed to random probability" (line 243). I could not make any sense of this.

Response: The authors have carefully revised these sections and added explanatory text regarding the concept of quantiles. In Section 3.4, we have included the definition of the probability ratio calculation formula and incorporated the use of standard deviation (SD) to characterize the variability of probability changes. Since quantiles and SD are commonly used statistical metrics in hydrometeorological studies, the authors did not provide extensive elaboration on them. We thank the reviewer for pointing out these issues, which have helped the authors further refine and improve the quality of this study.

Other comments:
Line 110: "The threshold for future periods is determined based on the threshold of historical period" is unclear to me. Does it mean they are equal? If not, what is formula to go from the threshold in the past to that in the futyre?

Response: We applied quantile delta mapping (QDM) to bias-correct the future precipitation and temperature data, ensuring that the thresholds for extreme precipitation and heatwave events, derived from reanalysis data during the historical period, can be consistently applied. In this study, we used the 90th percentile as the threshold to ensure a sufficient number of extreme events are captured. We have revised this sentence to improve clarity and accuracy. We thank the reviewer for highlighting this issue, which has helped the authors further refine and enhance the quality of this study.

Caption of Figure 1: "Identification ...events"

Response: The authors have revised Figure 1 to make its content more comprehensive and have also updated the figure caption. The authors thank the reviewer for pointing out this issue, as it helps us continuously improve the content and structure of the manuscript.

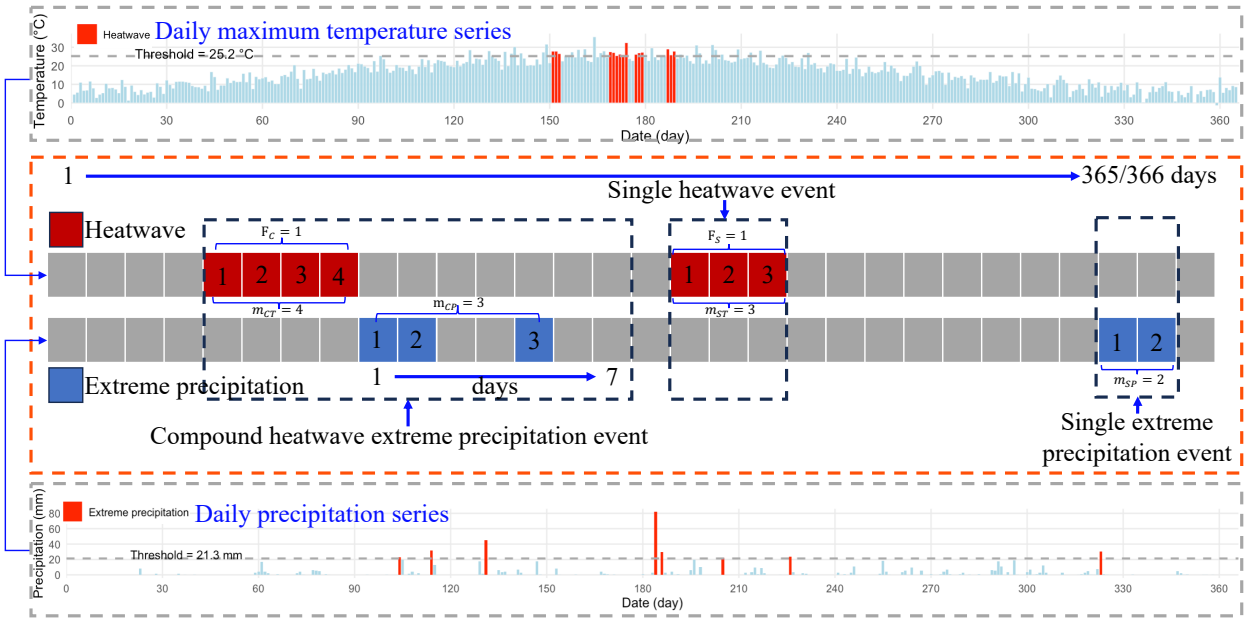

Figure 1. Schematic illustration of CHWEP and single extreme precipitation and heatwave events extracted from daily precipitation and daily maximum temperature time series over a year.

Line 167: One could argue that single heat waves are prevalent in all desert regions, where it only rains when it gets cooler

Response: As shown in Figure 2b, single heatwave events occur not only more frequently in desert regions but also exhibit high occurrence frequencies in tropical areas such as eastern Brazil and the Congo Basin, as well temperate regions including the U.S. West Coast and Central Asia. Furthermore, we employed a relative threshold to identify heatwave events in each region, which better captures the local characteristics of temperature variability. The authors have revised this sentence to ensure the description aligns more accurately with the results. We thank the reviewer for highlighting this issue, as it has helped the authors gain a deeper understanding of the significance and findings of this study.

Caption of all Figures: change "frequency" to "count"

Response: According to Equations 4 and 5 in Section 3.2, we calculated the frequencies of CHWEP and single extreme events in every year. Therefore, the authors have changed the label in all figures from “Count” to “Frequency,” which better aligns with the terminology commonly used in hydrometeorological extreme event studies. We thank the reviewer for pointing out this issue, as it helps improve the accuracy of the manuscript’s presentation.

Line 242, how are the standard deviations computed?

Response: The authors have already provided the formula for standard deviation (SD) in Equation 15 of Section 3.4. In this study, we calculated the SD of the observed encounter probability, the randomized encounter probability, and their ratio to quantify the interannual variability of CHWEP events. We thank the reviewer for raising this point, as it helps the authors continuously refine and enhance the content and structure of this study.

Caption of Figure 7: reference to panel (c) is missing

Response: The authors have added a reference to panel (c) in the caption of Figure 7. We have also carefully reviewed the entire manuscript to identify and correct similar oversights. The authors thank the reviewer for pointing out this omission, as it helps them continuously improve and refine this study.