
Response to the RC2

Dear Reviewer,

We sincerely thank the reviewer for your continued interest in and support of our research. Your insightful comments and constructive suggestions have been immensely helpful in further refining and enhancing both the content and structure of this study. Upon receiving your review, we immediately revised the manuscript in accordance with your recommendations, making every effort to meet all the revision requirements. The modified sections have been highlighted in red in the manuscript for ease of review. Below, we provide our point-by-point responses to your comments.

Thank you for your time and kind assistance.

Best regards,
Haoyu Jin

This study analyzes historical and future changes in compound heatwave–extreme precipitation events (CHWEPs) and compares them with single extreme events at the global scale from 1980–2100. They used three reanalysis datasets as observations and four CMIP6 models under two SSP scenarios (SSP2-4.5 and SSP5-8.5). They found that CHWEPs exhibit higher frequency, stronger precipitation, and longer heatwave duration in mid-to-high latitudes, and that CHWEPs occurring in the tropics feature more intense heatwaves than single heatwave events. They also suggest that these patterns will persist in future projections. In addition, they found that although both CHWEPs and single extreme events increase across most of the globe in the future, extreme precipitation following heatwaves will become notably more intense and frequent.

I agree that this is a very important topic in the field, but I find the manuscript still quite far from the standard typically expected for HESS. The choices of observational and model data are not reasonably justified, the statistical techniques are not analyzed in a way that convinces me of the robustness of the results, and the results/discussion remain largely descriptive and shallow (not quantitatively or statistically convincing). Many of the figures repeat similar information without adding new insights, and some key quantities are not clearly defined. In my view, the authors will need extensive methodological revision, additional testing, and substantial rewriting before the paper can be considered for publication in HESS. Below I list some of my concerns for the authors to address (with a few general examples):

The rationale for choosing ERA5, MERRA-2, and JRA-55 for observations is unclear. JRA-55 is also an older dataset, and it has already been replaced by JRA-3Q.

Response: The authors have added an explanation for selecting these three reanalysis datasets. First, the ERA-Land post-processed daily statistics provide daily precipitation and daily maximum temperature data that have been widely validated in numerous studies and are known for their high accuracy. Building on this, we further incorporated MERRA-2 and JRA-55 datasets. By averaging

these three reanalysis datasets, we constructed a more robust dataset, which serves as the basis for identifying extreme precipitation and heatwave events, and subsequently detecting CHWEP events. The inclusion of JRA-55 was also motivated by its relatively smaller data volume while still meeting the study's requirements, in future work, we plan to adopt the newer JRA-3Q dataset. We thank the reviewer for raising this point, as it helps us further refine and enhance the quality of this study.

Line 91: It is unclear and confusing what "ensemble mean" for observational data refers to here. The three reanalysis datasets are deterministic and have different systematic biases, so combining them as an ensemble requires a more explicit explanation.

Response: The authors have carefully revised this sentence. In this study, we selected three commonly used reanalysis datasets. Considering that reanalysis data may contain biases, we used the mean of these three datasets as the most reliable estimate. Based on this averaged time series, we derived thresholds for extreme precipitation and heatwaves, which were then used to identify single extreme events and CHWEP events. Similar to most climate change studies, we employed a multi-dataset mean to minimize potential biases associated with any single dataset. We thank the reviewer for pointing out this issue, as it helps the authors continuously improve the content and structure of this study and enhance the clarity of the manuscript.

The paper uses only four GCMs from CMIP6 without explaining why they were selected over others.

Response: In this study, the four selected GCMs offer higher spatial resolution and greater accuracy compared to other GCMs. More importantly, they provide daily maximum temperature data for both the historical period and the future SSP2-4.5 and SSP5-8.5 scenarios. Since heatwave events in this study are identified based on daily maximum temperature, this capability is crucial, many other GCMs rarely offer daily-resolution maximum temperature datasets. The authors have already included an explanation at this location. We sincerely thank the reviewer for their valuable comments and suggestions, which help us continually improve the content and structure of this manuscript.

In Section 3.1 (Identification and characteristics of CHWEP events), when defining heatwaves and extreme precipitation events, it would help to mention the actual threshold values being used.

Response: The authors have already added a description in Section 3.1 regarding how the thresholds for heatwave and extreme precipitation events were determined. Indeed, defining these thresholds is crucial. In this study, the threshold for heatwave events is based on the 90th percentile of the historical (1980-2020) gridded daily maximum temperature, while the threshold for extreme precipitation events is derived from the 90th percentile of the gridded precipitation data on wet days (>1 mm) during the same historical period. The 90th percentile was chosen to ensure a sufficient number of heatwave and extreme precipitation events for analysis. The authors appreciate the reviewer's comment on this point, as it helps them further refine and improve the study.

The definitions of formulas (1)–(8) are not very clear. Some metrics are calculated across all events while others are computed per event, and this is confusing. These formulas are difficult to interpret and need clearer and more accurate notation and description.

Response: The authors have revised the equations in Section 3.2 to enhance their clarity and added annotations for the relevant variables in Figure 1. Identifying single extreme events, as well as heatwaves, extreme precipitation events, and CHWEP (compound heatwave and extreme precipitation) events, is central to this study. Although these concepts are straightforward and intuitively understandable, their practical implementation demands substantial computational resources. In this study, for instance, analyzing single extreme events and CHWEP over the historical period (1980-2024) on a computer equipped with 8 GB of RAM and a 1.4 GHz Quad-Core Intel Core i5 processor required approximately nine hours. Given the differences among single extreme events and CHWEP events, all attributes of these events (except for occurrence frequency and duration) were averaged on a per-day basis in this study to facilitate comparison. We thank the reviewer for raising this point, which has helped the authors further refine and improve the work.

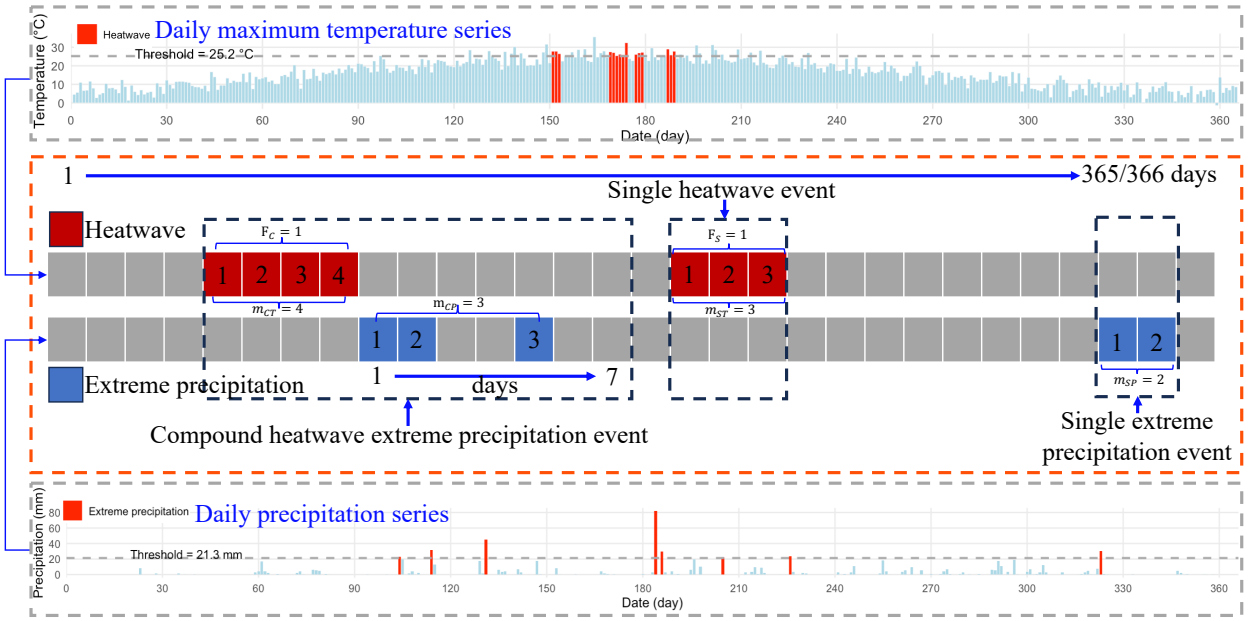

Figure 1. Schematic illustration of CHWEP and single extreme precipitation and heatwave events extracted from daily precipitation and daily maximum temperature time series over a year.

Some notations and units in the figures are not clear. For example, in Fig. 2a/b, what is the meaning of the units of "time"? Do you mean the total number of events over 1980–2024 or the number of events per year?

Response: The authors have revised the notations in Figures 2a and 2b to clarify that they represent the average annual number, namely the average frequency of single heatwave events and CHWEP events during the historical period (1980-2024). We thank the reviewer for pointing out this issue, which has helped the authors further refine the manuscript and enhance the accuracy of its presentation.

Some discussions of the results are insufficient or inaccurate. For example, in the discussion of Fig. 3c at line 181 ("Across most mid-to-high latitude areas, the extreme precipitation intensity during CHWEPs exceeds that of single extreme precipitation events"), the authors only focus on the Northern Hemisphere and ignore the Southern Hemisphere mid-high latitudes.

Response: The authors have carefully revised this statement to clarify that the significant regions are primarily concentrated in the mid- to high-latitude areas of the Northern Hemisphere. Although some parts of the mid- to high-latitude Southern Hemisphere also exhibit stronger precipitation in CHWEP events compared to single extreme precipitation events, the signal is not as pronounced. We thank the reviewer for highlighting this issue, which has helped the authors improve the accuracy of the study's presentation.

Much of the Results section is purely descriptive. The analysis could be more quantitative, for example by providing regional averages or more meaningful values to make the study deeper and more informative.

Response: The authors have thoroughly revised the Results section, providing a more in-depth description of the findings and expanding the comparison of regional differences to better understand how single extreme events, including heatwaves and extreme precipitation events, as well as CHWEP events, vary across regions. Overall, by comparing CHWEP events with single extreme events, this study has uncovered several intriguing phenomena: for instance, extreme precipitation following heatwaves tends to be more intense than precipitation from isolated extreme precipitation events; the mid- to high-latitude regions of the Northern Hemisphere emerge as hotspots for CHWEP occurrence; and extreme precipitation is more likely to occur after heatwave events. We thank the reviewer for raising this point, which has helped the authors further refine and enhance the quality of this study.

Some figures need to be modified and improved. For example, the colors for "Future SSP2-4.5 single" and "Future SSP5-8.5 single" in Fig. 5 are too similar to distinguish clearly.

Response: The authors have modified the line colors in Figure 5 to enhance their distinguishability. Similarly, the line colors in Figure 6 have also been adjusted for improved clarity. Furthermore, the authors have reviewed all figures throughout the manuscript and made selective color adjustments to enhance the visual clarity and prominence of the results. We thank the reviewer for pointing out these issues, which have helped the authors continuously improve the quality of the figures in this paper.

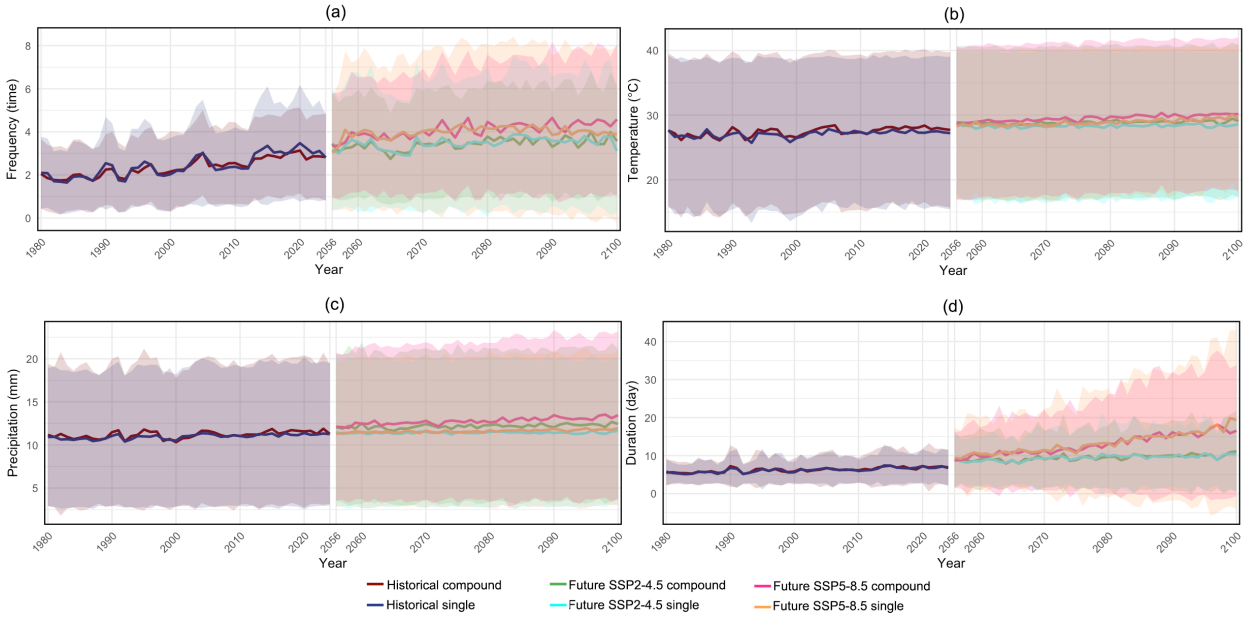

Figure 5. Changes in the frequency (a), heatwave intensity (b), extreme precipitation intensity (c), and heatwave duration (d) during the historical and future periods of CHWEP and single extreme events.

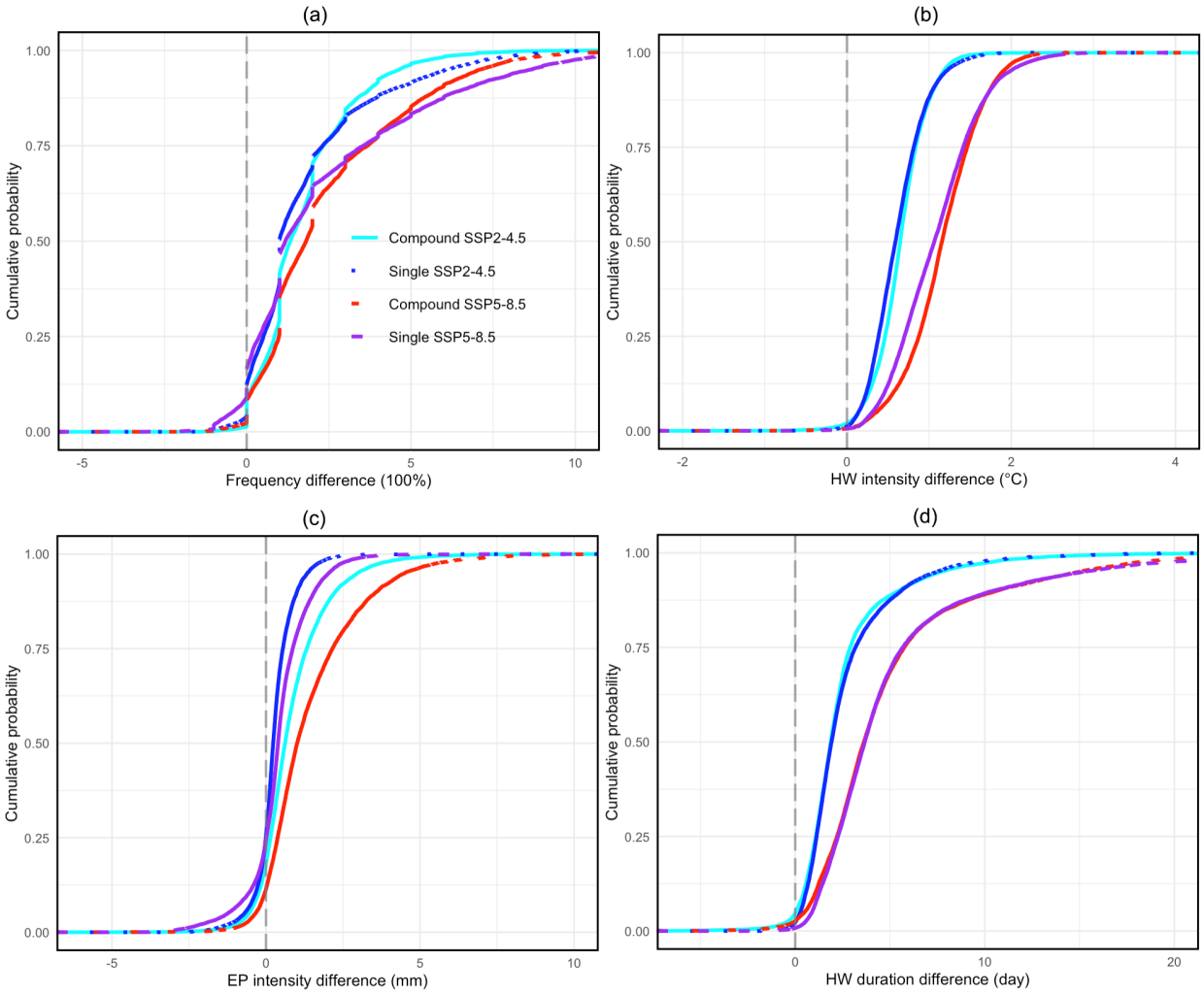

Figure 6. Comparison of the cumulative probability of the median differences between the frequency (a), heatwave intensity (b), extreme precipitation intensity (c), and heatwave duration (d) of CHWEP and single extreme events under the SSP2-4.5 and SSP5-8.5 scenarios and the historical period.

Although the authors introduce the Wilcoxon rank-sum test and event coincidence analysis in their Methods section, no p-values or other statistical results are shown. Without reporting these outcomes, the reader cannot assess whether the stated differences or assumptions are statistically meaningful.

Response: The authors have marked the regions where the median differences in the sequences obtained through the Wilcoxon rank-sum test (WRST) are statistically significant, and have further calculated the proportions of non-significant positive (NSP), significant positive (SP), non-significant negative (NSN), and significant negative (SN) results. We thank the reviewer for raising this point, which has helped the authors further refine and enhance the quality of this study.