

## Author's Response

We thank both referees for their review of our manuscript and the constructive comments that helped us to improve it. Below we address the request for technical corrections and the subsequent changes to the document.

1. *Line 12: The statement “Precipitation indices tend to emerge much later and mostly only later” would benefit from greater specificity. If possible, please indicate an approximate year or decade of emergence, as is done for temperature-related indices.*

We have added that precipitation indices tend to emerge mostly after 2030.

2. *Line 15: The abstract would benefit from a clearer closing statement. I recommend adding a brief sentence highlighting the broader implications of the results and/or providing a short outlook on potential applications or future research directions.*

We have added the following sentence at the end of the abstract:

“The results of this study provide a consistent basis for understanding and comparing the emergence of different types of extremes, and highlight opportunities for further research into the underlying drivers, as well as impact- or region-specific risk assessments.”

3. *Line 58: It is not sufficiently clear that the data were retrieved from the Copernicus Climate Data Store. Please make this explicit and state clearly whether the climate extreme indices were obtained directly from the Copernicus platform (e.g., “We retrieve the climate extreme indices from the Copernicus Climate Data Store ...”).*

We have clarified this and amended the paragraph as follows:

“In a previous project (Sandstad et al., 2022), ETCCDI indices were calculated for historical and future climate simulations from a large number of global climate models that are part of the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016). Some ETCCDI indices are percentile-based, and require a base period of time to define a reference climatology (here 1981–2010). The complete dataset of indices

was subsequently published on the Copernicus Climate Change Services Climate Data Store and is openly available (Sandstad et al., 2022).”

4. *Line 75-76: The statement made in this section requires clarification. Please specify the source of this claim or explain how it is demonstrated or supported by the analysis presented.*

This sentence was rephrased to “This corresponds to the point in time where we can start recognizing changes in climate extremes that exceed natural climate variability, as represented in the climate models (Hawkins and Sutton, 2012).”

5. *Line 145: It would be helpful to briefly summarize the approach of Brunner et al. (2020b), particularly explaining why certain models receive a larger weight and how this affects the results.*

We have added the following text to this paragraph and also slightly amended Figure 2 to improve readability:

“For this work, we have chosen to employ the scheme described in Brunner et al. (2020b) to weight all models comprising the multi-model ensemble. Weights consist of two components, the first being based on the individual model’s past performance for a target variable (here mean temperature or precipitation), compared to ERA5 reanalysis data (Hersbach et al., 2020). The second component measures a model’s level of independence from the rest of the ensemble, determined by a clustering of shared biases in climatology of surface air temperature and sea level pressure.

[...]

Figure 2 shows the weights given to each individual CMIP6 model used in this study. Note that several models receiving a low overall weight (e.g., IPSL-CM6A-LR, both MIROC models for temperature) also receive a low performance weight in Brunner et al. (2020b), while e.g., MPI-ESM1.2-HR and GFDL-ESM4 receive a high performance and overall weight in both studies. The CanESM5 model has a very high equilibrium climate sensitivity (Zelinka et al., 2020) and a large number of ensemble members (50), which together can lead to a multi-model ensemble that is biased towards faster warming if no weighting is applied (Hausfather et al., 2022). CanESM5 receiving a below-average weight, which is then distributed across all ensemble members, likely contributes to generally slightly later times of emergence in the weighted multi-model median, compared to the unweighted one.”