

## Review #2:

This paper presents a valuable and well-documented global bias-corrected and spatially downscaled SEAS5 seasonal forecast dataset. The dataset is likely to be useful for hydrological, agricultural, drought-monitoring, and climate-risk applications, particularly because it preserves the full ensemble structure over the hindcast and operational periods. I found the paper to be well-written, technically sound, and I appreciated the comprehensive analyses of the dataset.

I have only minor comments. Because the dataset is motivated in part by applications to droughts, heat waves, heavy precipitation, and other climate extremes, I recommend slightly expanding the discussion of performance for extremes. The paper already includes threshold-based Brier Skill Score analyses, including 10th and 90th percentile categories, but a more explicit synthesis of how reliable the product is for rare/extreme events would strengthen the manuscript.

Third, the statement that Empirical Quantile Mapping is well suited for precipitation extremes may require clarification. Standard EQM can have limitations for rare or out-of-sample extremes, particularly beyond the empirical calibration range. Additionally, the authors state they apply a linear extrapolation/scaling approach for precipitation extremes outside the historical CDF range, but a few more sentences explaining exactly how this works would be clarify understanding for readers.

Apart from these minor points and a few typographical issues (see attachment), I think this is a strong and useful contribution suitable for publication.

We thank the reviewer for this constructive comment and for highlighting the need for further clarification regarding the representation of extremes in the dataset and the behavior of the empirical quantile mapping (EQM) approach.

With respect to extreme events, we note that the evaluation already includes threshold-based probabilistic skill measures (including Brier Skill Score and ROC analysis for the 10th and 90th percentiles), which provide information on the reliability of forecasts for moderately rare conditions. We have now added a more explicit synthesis in the revised manuscript clarifying that forecast skill is generally higher for moderate anomalies, while performance decreases for increasingly rare extremes, particularly in arid and highly variable regions. This reflects the inherent reduction in sample size and increased uncertainty in the tails of the distribution. Nevertheless, the ensemble-based formulation retains useful probabilistic information for extreme event likelihoods, even where deterministic accuracy is limited.

Regarding EQM, we have clarified that empirical quantile mapping improves the marginal calibration of the forecast distribution but does not guarantee improved performance for all tail-dependent or threshold-based metrics. In particular, standard EQM is constrained by the empirical calibration range and may exhibit reduced robustness for rare or out-of-sample extremes. To address this, the methodology includes a linear extrapolation (scaling) approach for precipitation and an additive delta correction for temperature outside the empirical distribution range, which preserves relative anomalies at the boundaries of the observed

distribution. We have expanded the methodological description in the revised manuscript to make this treatment of extremes more explicit.

Overall, both EQM and the evaluation framework should be interpreted in the context of a probabilistic, impact-oriented dataset rather than a purely deterministic forecast verification framework.