The paper describes a dataset of bias adjusted seasonal forecasts of temperature and precipitation at monthly timescale with lead times up to seven months. The paper is describing the methodology well, besides some minor remarks, and provide relevant analyses to present the forecast skill and remaining deficiencies. I find some sections poorly described and at times confusing and in need of reformulation. I add minor remarks below:

L107: How is the +/- 15 d time window applied at the end points?

We thank the reviewer for this interesting question. The treatment of the ±15-day moving window near the beginning and end of the annual cycle required special consideration and was discussed during manuscript preparation. To clarify this aspect, we have expanded the revised manuscript and added an explanation of how the window is handled at the start and end points. For completeness, the original discussion is reproduced below.

Choice of window size

The use of a time window when calculating the shape of the cumulative distribution functions (CDFs) is particularly important for the effective correction of extreme events. As outlined in the Methods section, the window size plays a critical role in the success of the bias correction. A window that is too narrow may not sufficiently represent the local climatology, especially if extreme events are rare and thus either over- or underrepresented. Conversely, overly broad windows may blend data from different seasons, introducing inconsistencies, such as incorporating wet-season characteristics into dry-season corrections. However, for region-specific applications or certain use cases, a broader window may be advantageous. For example, Sangelantoni et al (2018) employed a 91-day sliding window to better reflect the seasonal characteristics and extremes of Central Italy, a region with relatively smooth transitions between seasons due to its maritime climate. Such a wide window would be less appropriate in a global context, where many regions experience sharp seasonal contrasts, particularly in continental climates. On a global scale, Gergel et al. (2024) used a 31-day window, emphasizing the need to preserve the shape of the distribution's tails.

The 15-day window used in each direction in this study (resulting in a 31-day window in total) was chosen as a compromise. It aligns with previous work by e.g., Thrasher et al. (2012), Themeßl et al. (2012) and is supported by Bedia et al. (2018), who showed that a 31-day window provides sufficiently smooth transitions between days while still being narrow enough to avoid the confounding influence of model drift in seasonal forecasts.

The use of a moving window approach in bias correction requires special handling of the first and last forecast lead days, since the full window cannot be constructed due to the start and end dates of the available SEAS5 data. Several strategies exist to address this limitation. Taking the example of a January forecast, the correction of January 1st values could be approached using one of the following methods:

1. Calculate the SEAS5-CDF with 16 days (0+1+15) and the ERA5-CDF with 31 days (15+1+15).
2. Calculate both the SEAS5-CDF and ERA5-CDF with 16 days each.
3. Use the SEAS5 values from the preceding December forecast to extend the SEAS5 days to 31.

4. Use the ERA5 values from the preceding December reanalysis to extend the SEAS5 days to 31.
5. Use the January 1st value 16 times to obtain 31 SEAS5 days.

As the forecast proceeds, the available window gradually expands, for example, January 2nd can be corrected using a 17-day window, until the full 31-day window is available by January 16th. The same logic applies in reverse to the last forecasted days (e.g., lead days 200 to 215).

Each method carries specific drawbacks. Method 1 leads to mismatched CDF constructions between SEAS5 and ERA5. Lorenz et al. (2021) reported statistical inconsistencies with this approach due to spillovers from prior months, resulting in biases in ERA5 that are not present in SEAS5. Method 5 suffers from a similar issue: overrepresentation of a single day, which distorts the distribution.
Method 3, which involves supplementing with data from the previous SEAS5 forecast, is sensitive to month-to-month drift effects in the model, potentially introducing systematic errors. Furthermore, this approach relies on outdated forecasts that may not represent the actual forecast conditions at initialization.
Gergel et al. (2024) addressed similar edge-window issues by extending the data range using preceding days, corresponding roughly to Method 4. However, since their work was based on climate projections, it is questionable whether this method is transferable to the context of seasonal forecasting. In addition, using ERA5 values to extend SEAS5 distributions compromises the independence between the forecast and reference datasets; essentially, identical values appear in both CDFs, effectively canceling out their impact. Mathematically, this is equivalent to reducing the number of independent observations, aligning with Method 2. Although Method 2 reduces the sample size for the CDFs, and thus slightly limits the capacity to correct extreme values, it avoids the major drawbacks of the other options. Therefore, the global bias correction method presented in this study adopts Method 2 as a conservative yet consistent solution.

L124: Please explain how the extrapolation is performed. How is the linear or additive methods determined? Is it trained on part of the tail?

The description of the treatment of precipitation and temperature extremes has been expanded in the revised manuscript. In particular, we now provide additional details on the extrapolation procedure and include the corresponding mathematical formulations.

L125: The data set shows some issues for dry regions. I attach two figures from a single forecast, where it is seen that an area with precipitation is constant over all bias adjusted ensemble members (looking at file SEAS5_BCSD_tp_1993.nc). This is a recurring problem with quantile mapping approaches, and requires special attention. I note this, but for the forecasts of monthly mean conditions, it is likely of little consequence, but for forcing hydrological models it might have impacts. When studying the data, you'll notice that the pattern over Sudan/Tchad, and south Algeria is constant across the members, whereas the patterns elsewhere changes as expected.

This is a very interesting find and we thank the reviewer for discovering the anomaly. Upon further investigation, we found that the pattern is indeed highly similar across ensemble members, but not identical. The feature originates from a precipitation band extending from southwest to northeast across the Sahara, which is also present in the corresponding ERA5 data for this month, albeit with lower intensity, indicating that SEAS5 slightly overestimated the event.

As January is climatologically very dry in this region, with a mean monthly precipitation of only ~3 mm during 1940–2024, even a single precipitation event can dominate the monthly total. Although SEAS5 ensemble members are generated from perturbed initial conditions, a strong signal already present at the beginning of the forecast can lead to very similar monthly precipitation patterns across ensemble members, with only relatively small differences between them.

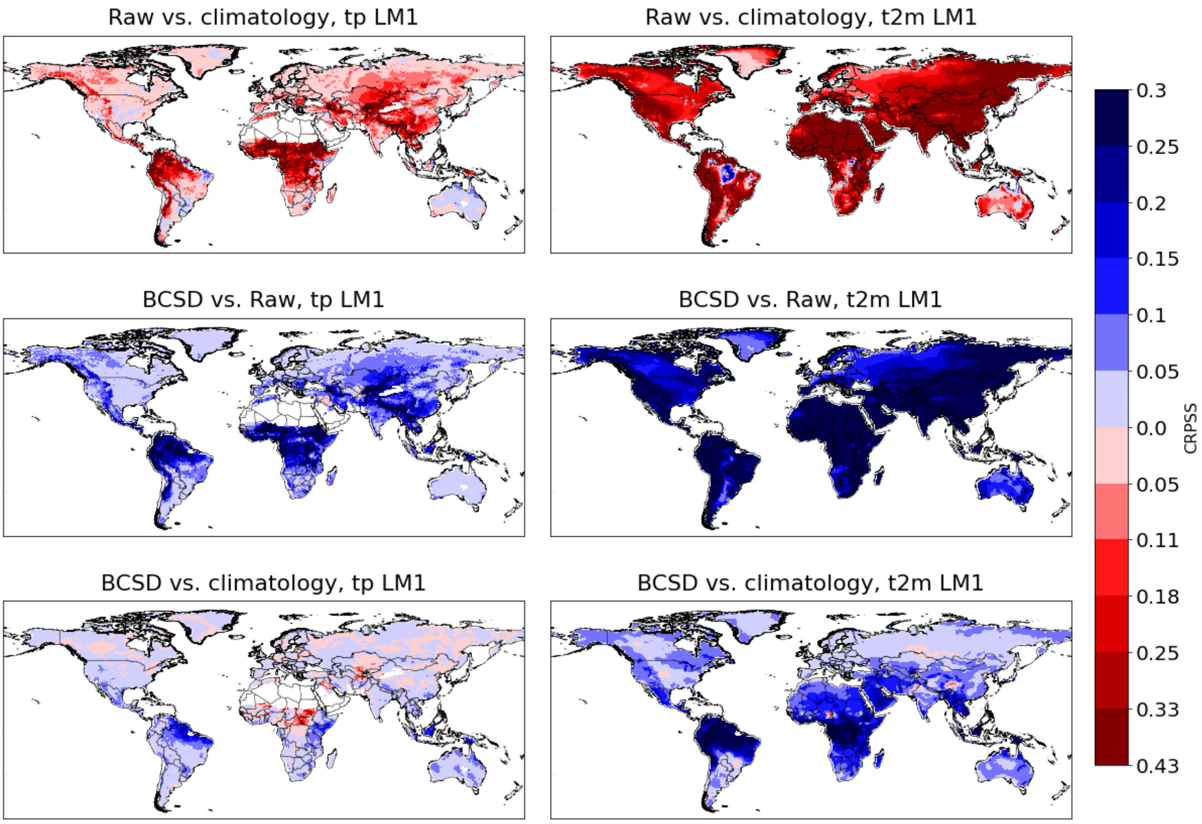
We therefore conclude that the apparent similarity does not result from the bias-correction procedure itself, but reflects a physically plausible forecast of a rare precipitation event in an otherwise extremely dry region. Examination of the underlying daily data confirms the existence of differences between ensemble members. If the reviewer is interested, we would be happy to provide access to the daily data via an anonymous THREDDS link.

L192-202: It is not clear if a particular case is shown, such as the bias in a forecast for a specific month, or if this is presenting the bias assessed from a longer time period of multiple years. L194 mentions “loss of skill”, but the figure shows bias and not skill, which should not be confused. Was temperature first “downscaled” by lapse rate correction, or is that included in the “bias”?

The bias is calculated as the absolute/relative deviation of the mean conditions from the reference period 1981 to 2016, using one month would of course not be robust. We clarified this in the caption of Fig. 1 and added it in text.

L205: “BCSD outperforms the uncorrected forecasts”, but only the BCSD result is shown. The uncorrected forecasts or the difference between the two should be shown to assess this statement. Figure 5 would better emphasize the deviations if values from 0.04-0.06 are shown, or simply the values minus 0.05 as deviations from the expected value.

The comparison between raw SEAS5 and SEAS5-BCSD is already illustrated in Fig. 1, which was designed to highlight the substantial biases present in the uncorrected forecasts. We therefore chose to focus Fig. 2 on the performance of the corrected product and to extend the analysis to additional lead months. Nevertheless, we agree that a direct comparison could provide additional insight (see figure below). If the reviewer and editor consider this useful, we would be happy to include the figure as supplementary material.



Regarding the scale of Fig. 5, we intentionally retained the full range of values to avoid overemphasizing relatively small deviations around the expected frequency.

L271: I find this sentence confusing, please reformulate. I do not see anything of linear behaviour in Fig 7, nor do I see any results for "share 33%". The use of "ensemble share" probably refers to the horizontal axis in the figure, but the figure also shown "Share of events", which I think adds to the confusion.

The sentence has been reformulated in the revised manuscript to improve clarity and now includes more explicit references to Fig. 7.