

**Editor decision:**

Dear Authors,

Thank you for submitting your revised manuscript, which has been read by myself and the same two original reviewers. All agree that the work is of high significance, quality and presentation. One reviewer has made an additional suggestion that is worth addressing, and will be re-reviewed by myself. I look forward to reading your final manuscript.

Yours,

Louise Slater

We are grateful for your kind decision. We have revised the manuscript to address the additional suggestion from Anonymous Referee #2. The responses are provided in the following.

Responses to Comments on “A statistical-dynamical approach for probabilistic prediction of sub-seasonal precipitation anomalies over 17 hydroclimatic regions in China” (Referee #2)

Anonymous Referee #2 Received and published on 2 November 2023.

Our responses are in blue and the revisions are underlined, with the reviewer’s comments shown as normal text.

**Suggestions for revision:**

Around L390 states that the relatively low reliability of merged forecasts at short lead times is due to the merged forecasts being too confident (no results shown). Later, around L445 it is discussed that the reliability of calibrated forecasts is poor. These statements require the reader to connect the dots. I suggest adding some text around L390 about the overconfidence of the calibration forecasts being the main cause of the merged forecasts becoming overconfident. It would be useful to refer to a figure in supplementary material, such as the quantile plots included in the response, or perhaps alpha-index plots for calibration & bridging. In this way, the two statements around L390 and L445 are linked and have evidence. Related to this, the abstract should be modified to say that some improvements to reliability are still needed at shorter lead times rather than reliability is high.

Thanks for your valuable suggestion.

We have added several sentences from **L. 374** to **L. 382** to provide evidence for the overconfidence at short lead times as follows:

To figure out the relatively low reliability at short lead times, we analyze the merged forecasts over Region 3 (Inland Rivers in Inner Mongolia) in May at a lead time of 0-day. The  $\alpha$ -index of merged forecasts is around 0.6, suggesting that the merged forecasts are of low reliability. We also investigate the model weights of calibrated forecasts and bridging forecasts. The results suggest that the calibrated forecasts are more important than bridging forecasts, which the cross-validated model weights are over 0.95. This suggests that the low reliability of merged forecasts is mostly caused by the low reliability of calibrated forecasts. Figure S9 presents the quantile ranges of calibrated forecasts and merged forecasts against time. The quantile ranges of both calibrated forecasts and merged forecasts are small, suggesting the forecasts are too narrow (too confident).

The quantile plots included in previous response are added as **supplementary material** in the revised manuscript as well.

The abstract has also been revised from **L. 23** to **L. 25** as follows:

Meanwhile, the BMA merged forecasts also show high reliability at longer lead times. However, some improvements to reliability are still needed at shorter lead times.

Typo:

L430: "do not always match"

Thanks for this comment. We have incorporated this suggestion in the revised manuscript.