Responses to Reviewer #1

Point #1

In the manuscript "Reconstructing climate trends adds skills to seasonal reference crop evapotranspiration forecasting", Yang et al adopted a new method to improve the prediction of evaporative water loss based on seasonal climate forecasts from the ECMWF model. This method is capable of dealing with the impacts of the changing climate on the prediction of future evapotranspiration (Reference crop evapotranspiration, ETo), and could lead to more realistic predictions. The changing climate has substantially altered the water cycle, representing one of the most critical challenges in hydrological modelling and water resource management. This work is innovative in taking this impact into account and addressing the challenges associated with climate change in the prediction of future evapotranspiration. The developed method is expected to be applicable to other models and thus benefit both forecasters (weather/climate centers) and forecast users (irrigators, hydrological modelers).

The manuscript is generally well written. Introduction clearly explains the background, challenges, motivation, and objective of this work; Method provides detailed information of the model, how the model runs are conducted, and evaluation metrics; Results generally are clear and readable; Discussion provides valuable insights and important implications for future improvements of climatology-based models in hydrological modeling and forecasting.

I encourage the authors to address the following issues before publishing this work.

Response: We appreciate the reviewer's nice summary and constructive comments.

Point #2

1. For time-series data, in addition to the magnitude of trend, another important feature is the statistical significance. I noticed the authors had taken this into consideration in selecting the months (8,9,10) for evaluating the performance of trend construction. In constructing the observed trends in calibrated forecasts, you empirically set limits of the trends in equation 8. I understand this is to avoid extremely large trend values. In addition to this adjustment, I think you should limit trends to zero, in grid cells where observed trends are insignificant ($P<0.05$). Otherwise, the trend reconstruction may overestimate climate trends. I see decreases in the correlation coefficients and skill scores when compared with the calibration without trend reconstruction (Figures 2 and 3). I think limiting the insignificant trends could avoid these unwanted decreases. I suggest the authors rerun the trend-reconstruction calibration and take statistical significance into account. If you see improvements in the new runs, update the results accordingly.

Response: We agree with the reviewer that the statistical significance of trends in observations should be tested and used to limit the reconstructed trends. We accepted your valuable suggestions and redid the calibration and analysis by setting limits in trend

reconstruction. Specifically, we used *P*<0.05 as the threshold to define statistically significant trends. For grid cells with insignificant observed trends (*P*>0.05), we set inferred trends to zero to avoid overfitting. We introduced this new strategy in section 2.3 as follows:

"For trends that are insignificant (P>0.05), we set m_i to 0 to avoid overfitting trends in calibrated forecasts. For significant trends, we set the m_i value based on trends in observations and raw forecasts during 1981-2019"

New results show that this strategy is not only effective in limiting the trend reconstruction to regions where observed trends are significant, but also helps avoid the reductions in correlation coefficient and CRPS skill score caused by overfitting (Figures 2 and 3):

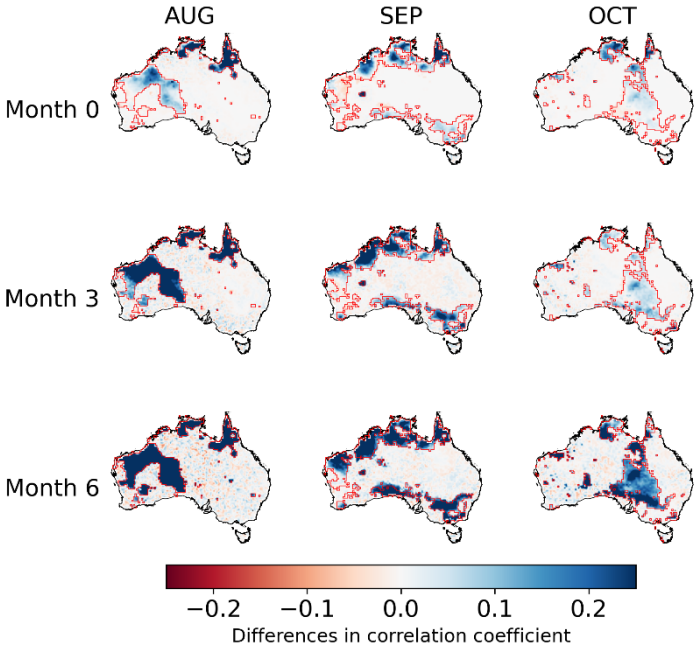


Figure 2. Differences in the correlation coefficient (*r*) between BJP-ti calibrated forecasts and observations with that between BJP calibrated forecasts and observations for three selected months (AUG, SEP, OCT) and three lead times (Months 0, 3, and 6). Red polygons show regions with significant trends.

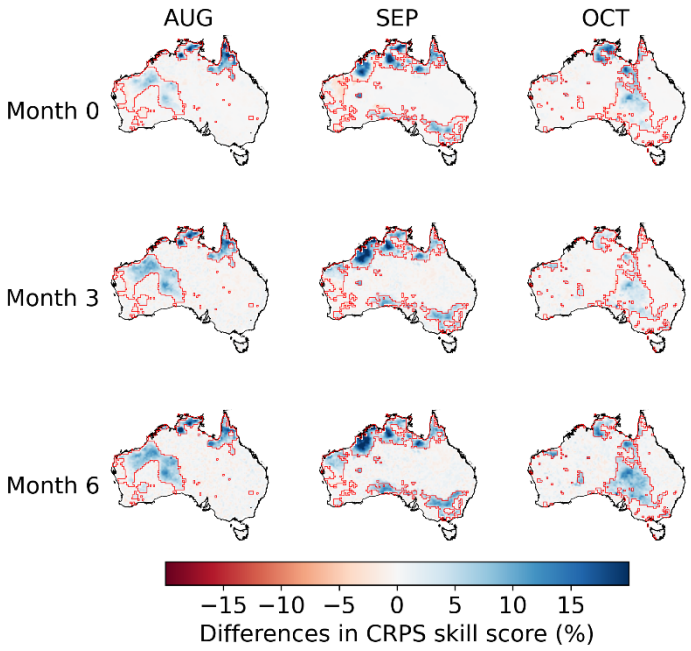


Figure 3. Differences in CRPS skill score between BJP-ti calibrated forecasts and the BJP calibrated forecasts for three selected months (AUG, SEP, OCT) and three lead times (Months 0, 3, and 6). Red polygons show regions with significant observed trends.

We updated all results in the manuscript based on the new calibration.

Point #3

2. In addition to the improvements in the 3 selected months, whether trend construction improve the calibration over the whole study period?

Response: Thank you for the valuable suggestions. We added a new figure (Figure 4) to show the overall improvements in CRPS skill score and updated section 3.3 accordingly:

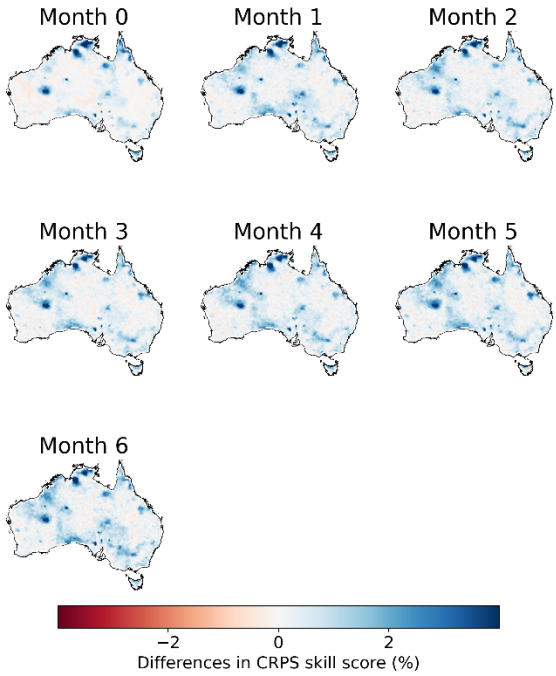
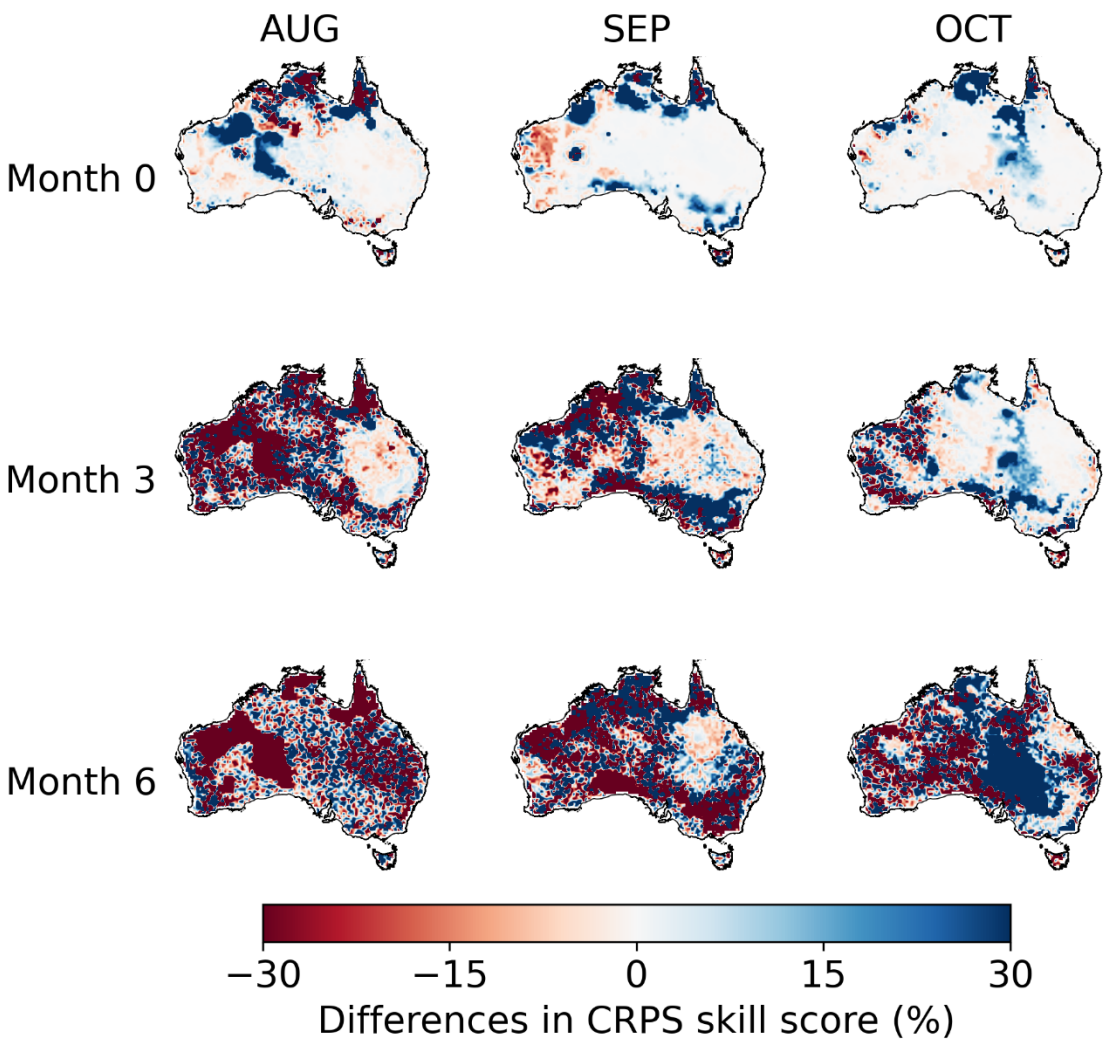


Figure 4. Differences in CRPS skill score between BJP-ti calibrated forecasts and the BJP calibrated forecasts over 1990-2019

Point #4

3. Presentation of the improvements in figures 2 and 3. I suggest the authors use the percentage of changes to demonstrate the differences. Since correlation and skill score vary largely from short to long lead times, using percentages could better demonstrate the more significant improvements at long lead times.

Response: Thank you for the valuable suggestions. We did not use percentage as the unit because we found that at long lead times, CRPS skill score in calibrated forecasts based on the BJP model could be slightly negative, and thus make the plot based on percentage confusing:



As a result, we decided to use their original unit. Actually, after fixing the problems in overfitting, figure 2 and 3 could better demonstrate how trend reconstruction improve the correlation and skill scores, particularly at long lead times. Please see details in our response to your comment #2.

Point #5

Specific comments:

Page 1. line 22, forecast should be forecasting

Response: We changed the wording accordingly.

Point #6

Page 3. line 92-93. This study is performed across Australia only

Response: We add the following sentence to clarify the spatial extent of this investigation:

"While SEAS5 produces climate forecasts across the globe, the calibration in this study is performed across Australia only.**"**

Point #7

Page 4. line 100, Calculation of ETo observations and forecasts

Response: We change the subtitle accordingly.

Point #8

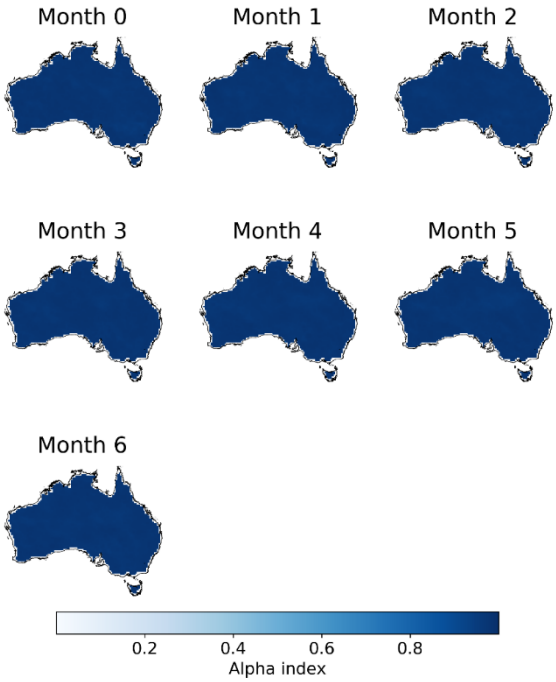
Page 6. line 160-165. Please italicize k in this paragraph and throughout the manuscript to be consistent with the equations.

Response: We italicized *k* in the manuscript.

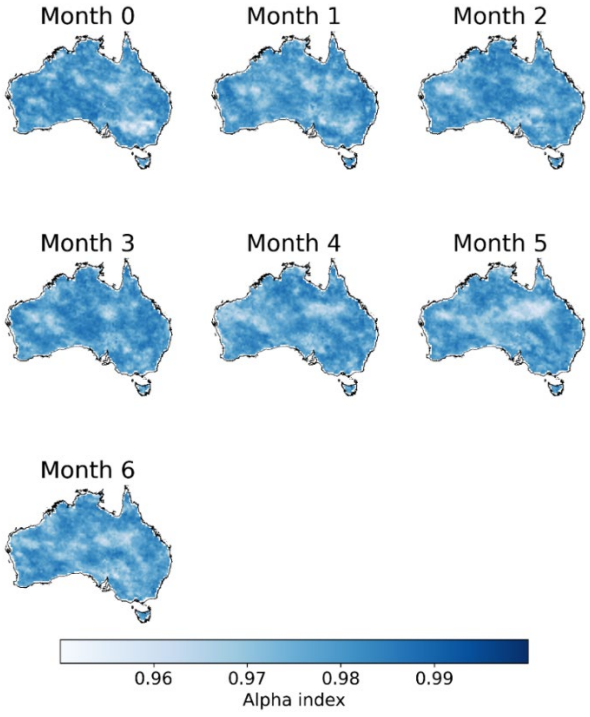
Point #9

Page 15. Figure 7, It is hard to read the alpha index values in the figure. Please consider changing the limits of the color bar, and use narrower limits (e.g.,0.8-1), to make the alpha index maps more readable.

Response: We replotted the figure with a new color bar of 0.95-1 and replaced the original figure:



With the following one:



Point #10

Page 17. line 378. To change with time?

Response: We changed the wording based on your suggestions.