

Responses to Reviewer 3

Summary

This paper proposes to increase the lead time at which operational forecasts of three month streamflow volumes can be issued by the Bureau of Meteorology in Australia. Currently, the data collection and processing cause the forecasts to be sometimes issued in the second week of the forecast period. The development of sub-seasonal forecasts in the region prompts the need for timelier forecasts. In this paper, the authors first present a forecasting method that would allow forecasts to be issued several days prior to the forecast period. Then, they investigate the relationship between forecast lead time and forecast quality. The authors show that being able to issue forecasts several days prior to the forecast period can be achieved without any loss in forecast reliability. Nevertheless, accuracy decreases as the lead time increases. The optimal trade-off between forecast accuracy and forecast lead time must then be decided on. Forecasts with lead times shorter than 7 days may offer an acceptable loss in accuracy as compared to the gain in anticipation time.

General comments

The topic is an interesting one and allows to look at forecast quality within an operational context, with timing constraints. The paper is very well written and is easy to follow. The proposed figures are also easy to read and well illustrate the paper. A catchment set is used to obtain relatively general results and the forecast evaluation criteria are relevant. The detailed comments are only minor suggestions and questions I had when reading the manuscript.

Detailed comments

Throughout the paper, “N-days lead time” and “N-month lead time” were confusing at first, probably due to the “N”. Could you maybe reformulate or add a small sentence in the first occurrences to define the terms?

Response: N is used to indicate any number of days or months. We can certainly make this clearer by using a different symbol or being more explicit.

In Section 2.3, have you compared the performance of the current forecasting system with that of the proposed system? Do you have an idea of their difference in skill, if the forecasts are both run at lead time 0?

Response: A very interesting question and one we pondered how to address when designing our study. Ultimately, we decided there are too many confounding factors to include a direct comparison in our paper.

However, in Australia, most of the forecast skill is acquired through the initial catchment condition predictors. Indeed, many models have no climate predictors. Therefore we expect the skill of the two systems to be relatively on par at 0 lead time.

Page 8, Lines 8-9: Is there a specific reason why you chose -5 and 5 as thresholds?

Response: These thresholds are essentially rules-of-thumb based on years of experience and some bootstrap experiments in previous studies. We can clarify this point.

Page 8, Lines 12-13: Could you maybe add a short sentence to further explain this? For instance, is the number of cases below -5 sufficient to assert this?

Response: The BJP modelling approach should theoretically produce a reliable climatology-like forecast in the absence of predictor-predictand relationships. In practice, under our stringent leave-five-years-out cross-validation, instances of skill negative skill can occur for various reasons related to short data records, low flows, a high density of zero flows, limited catchment memory etc.

We can certainly add a sentence or two to explain reasons for the small proportion of cases with negative skill. Additionally, we are able to include more information in the methodology section about BJP's propensity to produce climatology-like forecasts in the absence of a forecasting relationship.

Page 9, Lines 26-27: This sentence is, I think, very important in this paper, as it links forecast quality with operational expectations and requirements. I am just curious: have the authors further investigated the topic? How would you discuss / How have you discussed the acceptable loss in quality with operational managers? Is there a way to weigh the loss in skill more pragmatically, and relate it to the gain in anticipation time?

Response: The study was motivated by the demands of water managers to have timelier forecasts. The weather Bureau will continue to engage with operational managers to determine the best outcome for them. On pages 9-10, the benefit of having a flexible system is discussed, with regard to meeting the needs of operational managers.

Figure 4 and 5: It is interesting to put in parallel these figures with Figure 1 and Table 1. But how are the catchments ordered? I could not find it in the text, nor when looking at Table 1.

Response: The order of catchments in Table 1 and Figures 4 and 5 is not consistent – thanks for detecting. In the revision, we intend to order the catchments clockwise from northeast Australia to southwest Western Australia.

Figure 5: If I understand correctly, the objective of this figure is to show that the patterns between 0-day lead time and 7-days lead time are quite similar. Have you thought of showing the skill score of the 7-days lead time computed with the 0-day lead time as reference to highlight the differences between Figures 4 and 5? I understand that Figures 6 and 7 go in this direction but Page 8 Lines 3-6 could be illustrated in that way as well.

Response: It is certainly feasible to calculate the skill scores as suggested, to show the relative difference in error between the 0-day and 7-day lead time forecasts. However, as rightly pointed out, the differences in skill are illustrated by subsequent figures. Our preference is to leave the figures unchanged as they are effectively conveying the messages written into the text.