

Throughout this response, the reviewer's text is presented in black, our response in blue

This is an interesting and timely study into the value of forecasts for improving the performance of a simple water supply reservoir system with an operational trade off between augmentation of stored water through pumping and associated energy cost. The selected case study is appropriately simple and also informative for this type of analysis. Results are quite difficult to follow and key details are omitted from the method. The set of forecasts selected for use in the simulation are also poorly justified. Finally, I feel that the paper attempts to answer too many questions and would benefit significantly from more focus. For example, the analysis of the dynamical forecast product and its failure to provide skill over ESP is an interesting study in its own right, demanding much more in-depth analysis and interpretation than is offered in the paper. The operational section then addresses ESP vs dynamical and the additional question relating to importance of incorporating ensemble uncertainty. The paper would be much stronger if you were to focus on just one of these areas and deliver a more compelling conclusion backed up with in-depth analysis of a specific question. I recommend that the paper would be publishable if significant changes are made to simplify the overall story and provide further method detail as outlined in the comments below.

We thank the reviewer for their overall positive evaluation of our manuscript and the suggestions for improvement. We think the analysis of different scenarios is interesting (and other reviewers also seem to agree) and so we intend to keep it. Nevertheless, we appreciate that the manuscript writing is sometimes overly complex and that some analyses (for instance Fig. 6) raise more questions than they answer, so in the revised manuscript we will simplify some aspects of the Results section and Figure 4, while also editing the manuscript throughout for more clarity.

- It's not clear what optimization framework is used to deal with the forecast ensemble. The deterministic approach using rolling horizon (e.g., ANGHILERI et al., 2016) is quite common and there are very few successful examples in the literature where the full ensemble is used to inform the decision. Please outline exactly how the ensemble is used in your optimization and then justify the approach. If this is a new approach it perhaps needs to be described in its own, separate publication.

In our optimisation framework we minimise the expected values of the two objective functions based on the 51 ensemble members (whereas Anghileri et al. 2016 optimised the objective functions evaluated at the ensemble mean). In the revised manuscript, we will give a clearer explanation of this in Sec. 2.3.3. Regarding the justification of this approach, as we mention on Lines 424-431, the use of the full ensemble has been proved to improve the forecast value in several studies with shorter lead time than seasonal, while deterministic approaches such as the one applied by Anghileri et al (2016) did not show significant value in seasonal forecasts. So in our work we used the full ensemble at seasonal scale and also analysed the impact of the ensemble size, which confirm the value of using the full ensemble (see Section 3.2.2 and Figure 5): as discussed in lines 425-426, RTOS outperforms the current operation when using the ensemble forecasts, but it does not if uncertainty is removed and the ensemble mean is used. To further clarify this, we will add as an appendix the equations of the model and of the optimisation problem.

- Given the skill scores achieved for the dynamical forecasts, it's not clear why these were pursued in the operational part of the study. What is the justification for using a forecast product that is demonstrated to be unskillful relative to ESP?

One of the objectives of our work was to explore the skill-value relationship and whether one can extract value from forecasts in support of water resource management even if their skill is still relatively low. This is what led us to pursue the evaluation of DSP forecasts as well as ESP. As we mention in the conclusions (lines 480-482), this study indeed contribute to show that seasonal forecasts can deliver benefits to inform operational decisions even if their skill is low (as often the case in extra-tropical areas, such as the UK), and that under certain scenarios DSP can provide higher value than ESP despite its relatively similar skill.

- I found the results quite difficult to follow, partly because it's hard to keep track of the various operational settings. Why not simplify by showing the Pareto front for each forecast set (as opposed to five schemes with different symbols/colors). This would be both more comprehensive and easier to understand. Also, the emojis in the key figures are not appropriate.

We agree with the reviewer that our Figures are quite dense; on the other hand, showing each Pareto front in a different plot may make comparison across sets more difficult. In preparing our revised manuscript, we will try to find a clearer way to represent Figure 4.