Authors' Response

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

Report #1

I want to thank the authors for addressing previous comments and for their comprehensive replies. I found all replies satisfactory and the changes made to the manuscript significantly improve the quality of the paper. Removing Figure 6 and focusing the analysis of the value in Figure 7 answers my prior concern about the mismatch in terms of benchmarks. I list some additional minor comments below, including a methodological point (point 3) which could potentially be major.

REPLY: We thank again the reviewer for their overall positive evaluation of our manuscript and the suggestions for improvement.

L255 "based on dynamic weather forecasts": I suggest replacing this with "and the added performance of dynamic weather forcings rather than historical ones"

REPLY: We have replaced the original sentence by "and the added performance of dynamic weather forecasts".

L259: In the mean error equation, the averages over the number of members and time steps seem erroneous. The sums from 0 to M and from 0 to T should either be from 1 to M and 1 to T or the divisions on T+1 and M+1.

REPLY: Thanks for picking this up. This has been corrected, now the sum starts at 1 and ends in T and M.

L280-281: Here, it may be that the sentence needs reformulation or it may simply be a matter of terminology; otherwise it seems that you are averaging forecasts of different lead times but same forecast horizon. 1 Jan-1 Apr is a 3-month lead forecast of April 1st, 1 Feb-1 Apr a 2-month lead forecast of April 1st and 1 Mar-1 Apr a 1-month lead forecast of April 1st. Therefore, averaging these will not give a CRPSS for a 3-month lead forecast. Instead, it gives an average performance of the forecasts with a similar forecast horizon (but with different lead times).

REPLY: We agree with the reviewer that this needs to be further clarified. The average skill (CRPSS) for a given time step is the average skill of all the forecasts used from that time to the end of the simulation period (April 1st). This aims to represent the average skills of the forecast that a reservoir operator would have available during the operation process. This also makes the comparison between skills and value (which is also averaged over the simulation horizon) more meaningful. We have further clarified this by revising Section 3.1 (lines 277-283) and Figure 3.

L683: Replace "Solid lines represents" with "Solid lines represent"

REPLY: Corrected accordingly

Report #2

I have some remaining clarifications on the optimization approach using ensemble members. I understand from the revision that the forecast ensemble is used in the following way: evaluate decision set against each ensemble member separately, compute objective function for each, take expected (mean) value of objective function; optimize decision to minimize this expected value. This approach is typically avoided in multi-stage settings because it fails to account for decision recourse. This is why multi-stage stochastic programming using scenarios trees has been developed in prior studies, e.g.,:
https://www.sciencedirect.com/science/article/pii/S1364815212002770
https://www.sciencedirect.com/science/article/pii/S0309170814001262
Here you appear to have taken a shortcut, yet still achieved reasonable results. Please place your approach in this context and offer some discussion on why your approach works despite neglecting decision recourse.

REPLY: We thank again the reviewer for these comments. We agree with the reviewer that using multi-stage stochastic programming we could achieve better optimisation results and hence an improvement in the forecast value. However, as the Reviewer also pointed out, the simpler approach used here already provides reasonable results. Given the low forecast skill in this case study, we believe that the improvement in the forecast value by a more sophisticate optimisation approach is likely to be modest and possibly not sufficient to justify the increase in computational cost and complexity. In the revised manuscript, we have added a clarification and discussion of this point in the Formulation of the optimization problem (Supplementary material) and in the Discussion (lines 465-466).

I also don't quite understand why you claim that the approach is superior to the ensemble mean. Results in Figure 5 seem to suggest that the deterministic approach lies along a Pareto front for the chosen objectives. From the perspective of energy savings, aren't the deterministic results better than ensemble? I feel that the mean approach and the ensemble approach need to be compared more rigorously.

REPLY: Figure 5 shows that the deterministic approach only improves one objective, Pumping energy cost savings, at the expenses of deteriorating the other objective, Increase of resource availability by 1 April, and this happens no matter the decision priority that we select. While this may be acceptable in scenarios that prioritise energy savings (pso and psp), in the scenarios where resource availability is optimised individually (rao) or prioritised (rap), the fact that this objective is worse than in the benchmark means that using deterministic forecast has effectively no value. We clarified this point in Sec. 3.2.2 (lines 333-338) and in Sec. 4 (lines 423-430).