Review of

Assessing the value of seasonal hydrological forecasts for improving water resources management: insights from a pilot application in the UK

Peñuela, Hutton and Pianosi

The main merit of this paper is the proposal of the methodology. The paper forms a valuable contribution to the methodology of quantifying the value of forecasts, here in terms of water availability and energy cost. Probably the methodology is more widely applicable. Generally, the paper is well written, although sentences tend to be too long and their structure could sometimes be made clearer by repeating some short words.

Unfortunately, the conclusions from this paper are not really valuable. The problem with the first two conclusions, namely 1) seasonal forecasts can increase value and 2) ESP is hard to beat, is that they are case specific, as acknowledged by the authors (line 470). The third conclusion (the relationship between forecast skill and value is complex) is a trivial one.

Below, there is a quite long list of main points, which the authors have to address in my opinion: information about the observations should be given (p1), any procedure based a scenario or forecasts with more inflow than in the worst-case scenario seems beneficial, e.g. taking the median of the historical years (p2), the methodology should be better explained (ps 3, 5 and 6), Mliters are not a valid unit (p4), there is an issue with the bias correction (p7), different processing for the benchmark and other forecasts is questionable (p8), it is strange that the value of the driest years with DSP and ESP processing is not almost equal to the benchmark (p10) and the first part of the discussion section should perhaps be removed (p9).

In my opinion should be published after making the suggested major revisions.

Main points

1) A section about observations (discharge and meteorological forcing) should be added.
2) One of the results of this paper is that by basing the operational procedure on the forecasts, less energy for pumping is used while ensuring similar water availability (statistically over the years), compared to basing operational procedures on the worst-case scenario (driest historical year). It is my impression that any operational procedure based on forecasts or scenarios with more inflow into the reservoirs than in the worst-case scenario leads to less pumping and similar water storage, provided the increases are realistic. The authors confirm this in lines 395-397 for the case of applying a bias correction, which increases the inflows to the reservoirs and hence increases the value of the forecasts. So, the worst-case scenario is possibly easy-to-beat by any scenario with more inflow. Somewhere in the paper (in the discussion section?) the following points need to be discussed. Can this effect on value of increasing the inflow be generalized? What is the value of the forecasts if the operators base their procedure on the scenario of the year with the median value of the historical inflows? I suggest making a calculation with such a scenario. By the way: are the calculations in the worst-case scenario deterministic?
3) I did not understand Section 2.1 – 1b and c. These paragraphs need to be rewritten. At this stage this paragraph is too abstract. Perhaps providing an example of each concept (operation objective function, optimizer, set of operational decisions) would help. Perhaps merging

Sections 2.1 and 2.3 helps. Moreover, after 1b the "set of operational decisions is determined", so why is the operator again "selecting a set of optimal decisions" in 1c? Perhaps lines 124-126 helped me to understand a little bit of what you try to explain, namely that you use hindcasts to evaluate the performance of RTOS. If this is correct, just write that you use hindcasts to evaluate RTOS and discuss a possible operational application in the discussion section.

4) Replace all appearances of Ml and Ml per some time unit by m^3/s (per day is also ok), the common unit in the hydrological literature, e.g. in Figure 2, 3, 6 and 7.

5) I did not completely understand 2.3.1. Was river R fed by the outflow of reservoir S1 before the dam of S1 was built? Also, it sounds ridiculous to pump water, that was released by gravity from S1 to R, back to S1. So, is the water in R at the location where it is pumped out of the river partly fed by rivers that are not connected to S1? Is S1 located at lower elevation than D, so the water flow needs to be pumped?

6) I did not really understand those "rule curves" (lines 173-179). Add a figure with a rule curve. It is not clear to me how the refilling ($U_{R,S1}$) is done. Is the "missing water" immediately refilled or is the refilling spread over time until April 1, using the optimizer? In the latter case, how does the optimizer work? Can you give an example of an operational decision? What level is targeted on April 1? How does the operational procedure work for probabilistic forecasts? Since there is variation in resource availability by April 1 (e.g. in Figure 4), storage is not equal to the target on April 1. Can storage be larger than the target or is all water above the target spilled? Can storage be less than the target? Perhaps only in S2 and not in S1 because water can be pumped into the latter basin?

7) The method of bias correction is not correct (199-203). The number of years used to compute the multiplication factors differ per target year. I suggest using the common leave-one-year-out-method, i.e. the factor for each target year is computed from the data of all other years, including years later than the target year. Your method suggests that it is not allowed to use data from future years but there is no problem in doing so if different years are independent of each other.

8) Line 264 "but with three main variations": Why do you treat the benchmark differently? This implies that if the forecast is equal to the benchmark, the forecast value differs, which seems undesired.

9) The first general lesson in the discussion is "First, we found that the use of bias correction to improve the skill and value of DSP forecast is less straightforward than possibly expected" (lines 381-382). I do not agree. Such an expectation, namely that the forecast skill generally improves due to bias correction (is that the expectation?), just does not exist. Your study indeed confirms that this is a naive expectation. So, remove lines 378-393 or reformulate them. By the way: your bias corrections are based on observations of precipitation and temperature and not on the output (hydrological variables!) of ESP forecasts. So, I did not understand the sentences related to ESP in lines 387-388 and 391-392.

10) It is strange that the increase in the value of the system with DSP or ESP forecasts relative to the value of the system based on the worst-case scenario is highest in the driest years (e.g. lines 408-409), while those driest years resemble the worst-case scenario more than the other years. You need to explain this.

General text points

1) The authors often use the term “bias correction” without mentioning what is corrected. As far as I understood, the forcing of the hydrological model is corrected but if you do not repeat mentioning this now and then, it is confusing because the output of the hydrological model, i.e. inflow to the reservoirs, can also be bias-corrected. So, replace at numerous places “bias correction” by “bias correction of the meteorological forcing” or “bias correction of the forcing”.
2) In general sentences are too long, making the manuscript difficult to read. So, shorten sentences where there is an opportunity and make the structure of long sentences clearer, especially by adding some words in sentences with “and”. I made some suggestions below (e.g. 15-16 in the abstract).
3) “Uncertainty (considerations)” is used to discuss effects of ensemble size and the probabilistic nature of the forecasts. Replace throughout the paper the vague term “uncertainty (considerations)” by the more explicit terms “ensemble size” and “probabilistic/deterministic nature of the forecasts”.

Minor points

16	Insert “to” before “other factors”.
17	“Some of these factors” is too vague. Write which factors have a significant correlation with forecast value (see point Figure 6 below).
24	Add reference to endorse the statement that climate variability is increasing.
44-45	Replace “it provides” by “they provide” and add “that they” before “reflect”.
54	I miss the logic behind “i.e. ESP …”. Replace this part of the sentence or clarify the logic.
55-56	Reorder sentence to “The possible improvements of supply-hydropower systems operation due to the use of ESP were assessed by Alemu …”
69-71	Remove this sentence: this distracts too much.
78	weaker compared to ….?
83-92	Remove these sentences about some of the many existing metrics. It is not efficient to read about metrics not used in this paper and the metrics used in this evaluation are introduced 2.3.5.
94	Replace “this” by “the”.
99	Replace “simulate and compare” by “assess”. Simulate performance sounds strange and it is not clear from the rest of the sentence what is compared with what.
151	Insert “diagram” after “Pareto front”.
163	Consider removing all text about two companies. It is irrelevant for your story while it is making your story more complex.
168	Insert “(R)” after “river”
188	Remove period.

198 Did you also use a multiplicative factor for temperature?

211 $U_{S1,D}$ is also a pumped water flow according to Fig. 2.

214 Replace “The first objective function” by “Pumping savings” and “The second function” by “Resource availability”.

219 Replace “the 15%” by “only 15%”.

227 Rephrase sentence as follows: “They represent five different trade-offs of operational priorities, according to their relative importance”

237 Remove sentence.

247 Remove “and for a given lead time”. The role of lead time comes some sentences below.

251 Replace “lead time” by “range of the lead time (we use monthly ranges)” and “CRPS values” by “individual CRPS values”.

252 Replace “CRPS” by “individual CRPS values”.

253 I suggest to replace “mean error” by “discharge bias” since bias in the common word for mean error and the addition of “discharge” helps to distinguish this bias from that in the forcing. I also find the equation redundant. Just write that the bias is the difference between the means of the forecasts and the observation over all …….”

262 Add “(1975-76)” after “drought on records”.

Figure 3 Is this the sum of the inflows to both reservoirs? Are these results for the whole year or a specific part of the year? In the legend of the lower panel “2006” should be replaced by “2016”.

334 and 338 Replace “Figure 6” by “Figure 5”.

Figure 6 If I just look at these graphs, I get the impression that there is no significant relationship in any of these graphs. However, according to your p-values relationships are significant at the 90% confidence level in panels b and e. Is the calculation of the p-values correct? Or are those low p-values due to using the Spearman coefficient instead of the Pearson coefficient? I think you should use Pearson unless you have good reasons to use Spearman.

344 Replace “skill” now and then by “forecast skill”, to remember the reader what type of skill this is.

358 Replace “year” by “years”.

395 Replace “reduce” by “reduces”.

399 “improvement of forecast accuracy in some direction”. What do you mean by “accuracy”? For me this is something like the root-mean-square-error, which means that there is only one desired direction, namely towards 0. Do you mean something like “a change towards either higher or lower values can be more valuable than a change in the other direction”?

411 Is “Initial storage (total storage value)”equal to “Initial storage” in Figure 6? For clarity, be consequent in the use of specific terms. Moreover, panels c and h in Figure 6 do not show a significant correlation (p-values of 0.21 and 0.80).

428 Remove “)”

447 Replace, for clarity, “seasonal forecasts” by “seasonal meteorological forecasts”.

465 Insert “of” before “the institutional”

466 Insert “of” before “the most”

478-479 Replace “but also the methodology in the first place” by “but in the first place by the methodology”.