Interactive comment on “Analysis of the effects of biases in ESP forecasts on electricity production in hydropower reservoir management” by Richard Arsenault and Pascal Côté

Anonymous Referee #2

Received and published: 30 December 2018

This manuscript presents a study on the effects of bias in seasonal forecasts developed using the well-known Ensemble Streamflow Prediction (ESP) approach on release decisions from a series of reservoirs for the generation of hydroelectric power stations. The energy generated is destined primarily for use in the Aluminium smelting processes. This study presents a nice example of the use of seasonal hydrological forecasts in decision-making, relating probabilistic forecasts and their typically inherent biases to the decisions that are made with these forecasts.

This study is of interest to the readership of HESS, and although I think that the different methods used for the optimising the releases informed by the seasonal forecasts are relevant, I think that the main interest (as also suggested by the title) are in how uncertainties and biases influence the optimal decisions made. I would, however, suggest some improvements and clarifications to the manuscript to increase the noted appeal to the readership of HESS. One of the main results that the authors seem to conclude is that a forecast without bias is not necessarily as beneficial as when there the forecast has positive bias. To most hydrologists working in seasonal hydrological forecasting, this seems to go against what is often considered as the ultimate goal of bias correction methods: developing an unbiased forecasts. Although the authors elude to it to some extent, one of the main reasons for this being that this tends to avoid spill, which is penalised in the optimisation as the volume of water that is spilled is then not used in generating power.

This is particularly so when the minimum base load constraint is not included, as then the optimal solutions then tend to run the reservoirs at low heads (though this will generate less power for the same release discharge, and may incur higher penalties due to the recreation constraint). When the minimum base load constraint this changes, as this is now imposed as a constraint rather than being included as a penalty. Including this constraint reduces the “room” for the optimisation algorithm. I think that this discussion is interesting, but do think it should be generalised in the discussion. The conclusions found are not general to the use of probabilistic forecasts, but are conditional on the shape of the decision making problem (as formulated in the optimization function). This sheds an interesting light on the value of forecasts, and how value is related to the relative penalties imposed by the different parts of the objective functions (for example recreation versus hydropower generation).

I think it would be good if the discussion is extended to reflect how the conclusions found would change if the shape of the objective function changes. What would happen if the hydropower objectives changed (the current requirement would seem to favour a steady load, rather than for example hydro-peaking), how would this change if there were additional constraints or penalties on downstream releases (I would suspect a flood damage penalty would result in the same conclusions as this would also favour spillage being avoided, but an environmental constraint may favour spilling). I think the essence of my comment is that I agree with the authors that it would seem that a biased forecasts is to be preferred but this needs to be considered from the point of view of the decision process that the forecasts are used to inform. From the point of view of the hydrological forecast in its own right it make sense that the forecast is as unbiased as possible. That there is more value in the biased forecast in this case is in essence the result of a transfer of risk through an objective function that is not symmetrical. This risk transfer may work very differently in a different setting. Another good example is in water allocation from reservoirs for downstream irrigation. In this case it would be of value to tend towards a low bias in the ESP ensemble as this avoids the risk of insufficient resource being available to farmers who may have planted on the expectation of a higher availability of water. It may also be interesting to understand if the biases in the ESP forecasts have different effects through the season. I can imagine that the value of the ESP forecast, as well as the effect of biases differs depending on the time of year.

We would like to thank the reviewer for this comment. The analysis is 100% correct. Our work shows that the impact of bias in the forecasts is very dependent on the optimization problem (i.e. the response surface of the objective-function). We have modified the discussion to make it clearer that this is the case, and we have given a few examples that we think could lead to similar results. We added the following sentences in section 5.2:

In essence, the more flexibility the system has, the more it can try to leverage high head (which leads to higher performance but more spilling due to the lack of information about uncertainty). Any constraint or penalty that limits this flexibility by forcing lower reservoir levels (such as high costs of flooding, environmental flows or required base load generation) removes the option of over-filling the reservoir and thus leads to less spills. Therefore, the value of forecast quality is highly dependent on the objective function used in the management process. In a highly constrained system, the value of generating unbiased forecasts can be essentially zero. Conversely, a system with few constraints can make use of better forecasts but then the optimization method must also not introduce new biases as was shown in this paper.

In the paper the approach the authors take is to use the simulated historical discharges as the reference, effectively removing issues with model biases and errors. Though I agree that this is a sound approach, application of the approach in a real situation would mean that these biases and errors should be considered. It would add to the paper if this is discussed. What are the additional biases that would then need to be taken into consideration?

We have added a few sentences in section 5.1. Limitations:

In the same vein, the simplification of the system is reflected in the hydrological model, which would normally introduce new biases. The methodology used herein eliminates this bias, but it would need to be taken into account in a real-world application.

Specific comments: Page 2; Lines 5-10: This section discusses the usefulness of ESP forecasts, commenting that these are particularly of value in the longer term. I think this discussion should be qualified to some extent. The skill of ESP when using climatology as a reference is derived from the persistence of the initial states. For hydrological systems that have a strong nival regime, such as in the case study presented here, there may be such a persistence. This also implies that the skill varies seasonally, being highest at the start of the snowmelt season, and lowest at the start of the snow accumulation season. This seems to be eluded to later (page 8, line 15) where the near zero correlation between the model states and future inflows is mentioned. This would mean that at this point the ESP forecast is solely based on climatology, which means that there is no skill.

This is a good distinction, we have clarified it in the text:

The skill of ESP is largely based on the persistence of the initial states, which themselves depend on the dominating processes. In the case of snowmelt-dominated catchments, ESP forecasts can be relatively skillful at the beginning of the snowmelt period, whereas at the beginning of the snow accumulation phase, the initial states play essentially no role on long-term inflow forecasts meaning that the ESP skill is entirely based on the climatology (Harrigan et al., 2018).

Page 7; Line 30: The authors suggest in the discussion on considering the simulated discharge as forecast target that this by-passes issues related to the initial state of the model. I think it would be more appropriate to refer to this as bypassing issues with model error and how representative the initial state in the model is of the true hydrological conditions at the start of the forecast.

Thank you, this is a good point and we have added this detail in the sentence:

The historic simulated streamflow is considered as the forecast target, bypassing all issues related to the hydrological model's errors and its representation of the initial conditions as compared to the true hydrological conditions at the start of the forecast.

Page 8; line 17-20: The construction of the ESP forecast is discussed where the available data is divided into two periods, with the test-bed run for the second period using the ESP ensemble in the second period. Later in the paper the authors elude to possible issues as a results of the inhomogeneity between these two periods. I think it is relevant to explore this further. Is there a bias introduced due to the selection of the periods, and if so, what is the sign of this bias and how does this compare to the multiplier applied to the ESP to produce biased forecasts.

This is an excellent point that we have investigated further. We have added the following lines to section 5.1, 3rd paragraph:

In this study, average inflow volumes between the calibration period and simulation period over the entire system differ by less than 2% (1405 m^3/s in calibration, 1432 m^3/s in simulation). This could explain a small portion of the results, however the differences are mostly found during the snowmelt period during which there are spills (and maximum generation), limiting the impacts on the rest of the year where the optimization problem is more difficult. There is no doubt however that during other periods, the fact that there is more water in simulation than in calibration would point to positively biased forecast being more efficient.

Page 10; Line 21: I found the notation in the equations somewhat confusing. It is not so clear what K. T, J, pertain to. It would be useful if these could be explained.

We have added the lacking information. First, K is a mathtype translation error, it should read "…" in the series from "1, 2, …, J" for example, instead of "1, 2, K, J". This has been changed in the manuscript, so K is no longer an issue. T is the total number of time steps, and J is the total number of hyperplanes used to approximate the water-value function.

Page 11; Line 4: Unit outages are mentioned, but little detail is provided of what the implications of these outages are. To what extent do these restrict the available options for meeting the objectives? How are these outages determined? Is this according to a set schedule or according to historical data?

We have added the following sentence, referring to the set schedule of unit outages:
"These shortages were forced according to the average outages seen at the different powerhouses, either due to planned maintenance or unit failures. An average calendar was used due to the fact that maintenance outages are relatively simple to reschedule and compose the bulk of unit outages in this system."

Page 11; Line 26: “incorporates on branching” – this is a strange formulation; perhaps rephrase.

Thank you, this was a typo. It should have read “…incorporates only one branching…”. It has been changed in the manuscript.

Page 24; Figure 3: The scaling of the b) figure does not readily help its interpretation. Clearly the LSJ is the larger catchment and hence much larger inflows. Perhaps use relative volumes, with an indication of what the average volumes is in each. Is the flow in the left column the mean across the 3-month season? If so, then is this not the same as the volume (the only difference being the multiplication with time).

We have modified the figure to indicate relative inflows in the right column. We also indicated the average values in the figures to be able to reconstruct the absolute values. As for the left column, the units are percentages of relative bias, therefore the use of average inflow or total volume does not change the interpretation of the figure.