Response to Anonymous Referee #1

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 #1

Received and published: 16 August 2018

*** General comments ***

This is a very interesting paper studying, in detail, the effect of forecast bias on electricity production in hydropower reservoir management.

While each and every of the results presented is interesting, I am not convinced by the authors' analysis of the supposedly beneficial effect of a positive streamflow forecast bias on the generation output. While some bias appears to indeed be beneficial for this particular optimization model, it may not be beneficial in general. More on that below.

Thank you for these comments. Below we address the specific points raised by anonymous referee #1 and indicate how we updated the paper to reflect the changes. To be clear, we do not advocate aiming for biased forecasts; instead, we want to show that the value (or impacts) of forecast bias can change dramatically based on the hydropower system setup and constraints.

*** Specific comments ***

The comments in this section center on the generation output as a function of streamflow forecast bias, and on the results leading to Figure 7 in particular. First of all, some things are not clear to me:

- How is the relative MW ratio computed exactly? Is the result of an open-loop application of optimized reservoir releases to the simulation over the optimization horizon? Or is a closed-loop approach used to produce these figures, where re-optimization is performed at every simulation time step?

In our study, the optimization is a closed-loop system, as shown in figure 2. There is a re-optimization at each step, with the period's generated ESP forecasts. The decision for the first period is then used to simulate the state of the system for the next period, including reservoir levels, spills and generated energy along the system during the period. We then take the average MW generation of all periods and compare that to our baseline value to transform the absolute to relative MW (MW ratio).

- What is the impact of the choice of values for the parameters lambda and eta (Equation 1 on p10)? It seems to me that higher lambda values would also entail more conservative operation and would hence affect the results presented in Figure 7.

Yes, the weights play a role in the model behaviour, but not necessarily as stated above. A higher value of Lambda would penalize energy shortages, and because the reservoirs are finite and some inflow scenarios are quite dry, it is impossible to always respect this constraint. Maximum levels are hard constraints due to the physical assets that simply cannot be overtopped. Minimum levels are dictated by the water intakes. A higher Eta value, which penalizes constraint violations to summer environmental flows, would penalize water shortages and would thus encourage the model to keep more water reserves for the summer months.

In addition, given the fact that some constraints are impossible to respect 100% of the time, we can follow two options. The first is to give very high penalties to the violation of constraints and then measure how that impacts energy generation. The second method is to try and balance the penalties so the output of the simulation has similar statistics to the historical operation of the system. In our case, the parameter values were selected as they provided similar energy and reservoir levels (and constraint violations) as were observed in the historical dataset.

The authors point out (p16) the tendency of deterministic methods to be overconfident in their ability to manage a reservoir at high head, thereby causing larger spillage than necessary:

- This will indeed be an issue if the optimization results are applied in an open loop setting. However, if re-optimization is performed every simulation time step in a closed loop setting, the planning will adjust to higher-than-anticipated reservoir levels and spilling should be much reduced.

This is true indeed, and tests with an open-loop optimization show that the spills are much larger than what we see using the closed-loop optimization. However, the closed-loop optimization still uses the entire forecast duration to determine the best decision at the current time step. Even in a closed-loop system, the optimization model determines the best course of action given what it thinks is a perfect forecast and will then attempt to keep reservoir levels as high as possible, to later release the water at the last possible moment to maximize the duration of the higher efficiency turbinating. Then, when the future inflows are higher than anticipated at the next time step, the optimization model must spill the excess water but will only spill the amount it thinks is required to keep the maximum efficiency, which leads it to again be in a higher than desirable state. This all comes back to the fact that the deterministic methods do not consider uncertainty in their decisions, which leads to overconfident decision-making.

- Use of a soft upper reservoir water level constraint, rather than a hard constraint, would probably eliminate the spilling issue altogether (in a closed-loop setting).

As stated above, the upper water level constraint is a hard constraint as it is related to dam safety. We could model the system with some tolerance but this would go against the operational policy. When the limit is reached, there is a mandatory spilling operation to ensure the level does not rise further. These were all performed in a closed-loop setting.

- With the spillage issue out of the way, the reduced reservoir levels resulting from the positive bias should, in the long run, negatively impact generation output due to a) reduced head and hence reduced efficiency, and b) due to reduced water availability beyond the optimization horizon.

We understand the rationale, but we disagree with this statement. The spillage we see is the result of being overconfident in the future inflow forecasts, and as long as there is a gain to be made to creep closer to the reservoir maximum limits, the model will do so. Of course, if we penalize the reservoir limits too much, then the risk will be reduced but the model will be too conservative.

As a result, I am not convinced that the reduced spillage/higher generation output phenomenon is fundamental, and therefore I would suggest to be much more cautious in claiming that a small positive streamflow forecast bias is desirable (p18). Rather, it strikes me as a phenomenon that emerges out of the interaction between forecast bias and (perhaps, if I understood correctly) the lack of a closed loop, and too stringent reservoir level bound modelling.

We agree that the statement about having a forecast bias being desirable is misleading. What we meant is that in this case, we can factually demonstrate that a positive bias improves the results, due entirely to the fact that the optimization methods are imperfect. Therefore, we want to show that in these cases, the optimization methods are biased and that using an oppositely biased forecast can help correct the optimization model's shortcomings. Of course, we think everyone agrees that what we need to target as a community is having unbiased optimization methods with unbiased forecasts. With this paper, we want to draw attention to the fact that there is indeed a risk in using these methods which are imperfect.

Accordingly, we modified the paper to ensure that it reflects these thoughts. We had it read by an independent researcher to ensure that the message we are trying to convey was understood correctly.

*** Technical corrections ***

p7: The need to derive adequate hydrological model initial conditions is pointed out. Then, it is described that these are derived using a hydrological model driven by observed climate data. To me, this begs the question on how this model is then initialized before it is ran "once more until the forecast date"?
We start the model "empty", i.e. with no water in any of the water stores. After one year, the model spin-up is complete. When we perform a forecast, we can simply take the current day's initial states or re-run the model starting from day 1 (empty reservoirs) and let the model arrive to the current day, thus regenerating the model states on-the-fly. We changed the text accordingly.

p9: It would be helpful to include a formula describing how exactly the relative bias is computed.
Indeed, this is a good idea. We have added the equation as Eq. (1) in page 9 and updated the other equations accordingly.

p11: Equation 5. I don't see how the fundamentally nonconvex product of discharge Q and head H can be approximated using a set of linear inequalities; consider for example the relation QH restricted to Q=H, this is a convex function, which can – after approximation with a bundle of linear inequalities – be used as a lower bound for the power generation, but not an upper bound (due to the hyperplanes intersecting below the curve). The reverse holds for the relation QH restricted to Q=H_max - H, for example, which is concave. Not sure what the impact of the hyperplane approximation is on your results, but it looks like there will be issues with the head dependence of the power generation. Consider looking at some of the recent work on the homotopy approach towards tackling the QH nonlinearity without sacrificing physical accuracy.

First, thanks for the information about homotopy applied to this kind of problem. We will consider digging deeper in this aspect in future works, we hope you understand that we consider modifying our entire approach out of the scope of this study.

Second, the modelling of the Q-H is a convex piecewise linear approximation. We take the lower bound of the envelope of the hyperplanes to approximate the real Q-H-P surface, which is convex. It is important to note, however, that in our case study (and quite probably that this is not the case for all systems) the approximation errors of the efficiency are typically found in the very-low Q section of the surface, when Q rapidly transitions from low to high values. In those cases, there can be a small error on H due to the downstream levels affecting net head.. However, for the vast majority of the historical dataset, the operations fluctuate between max flow and approximately half of that. The most critical errors are those found in the high flow cases, but these are very well modelled in the system which uses non-linear functions to best approximate the change in efficiency during transitions in those ranges. Therefore, we are rarely in the lower bounds of the operating policy, which keeps us away from the “problematic” non-convex parts of the problem.

p16: I find referring to the scenario tree approach as being a deterministic approach confusing. Yes, the algorithm is deterministic, but it takes forecast uncertainty into account to some extent and is in that sense probabilistic.

Yes, we agree. We tried explaining that we are using deterministic methods in a stochastic setting, but this phrasing is not clear we modified the sentence to read:
While the algorithm is deterministic in that it returns the same response to the identical inputs, it does make use of multiple scenarios and in that sense, it can be seen as probabilistic.

In general, it is also not immediately clear that the "unique decision method" is the same as the "scenario tree approach". Best to make this explicit earlier on.

We agree, we changed the occurrences of “scenario tree” to “unique decision method”. We still use scenario fan to explain that there is a fan of scenarios, but not in the sense of the “scenario tree optimisation” method.

Figure 6: The units of panels (b) and (c) on the X axis don’t make sense to me, esp. the negative efficiencies.

The efficiency is measured as the Energy output per volume of turbinated flow, i.e. MW/(m3/s). The higher the number, the more energy is being produced for the same amount of water. The values in panels b) and c) of figure 6 are relative values, i.e. the value of the 80th percentile (for example) as compared to the value of the 50th percentile member. Therefore negative values indicate that those years performed less well (in terms of efficiency) than the median case, and positive values indicate better efficiency as compared to the median case. We changed the axis labels to indicate that it is in fact “Difference in average annual efficiency”. Thank you for catching this.