Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., author comment AC2
https://doi.org/10.5194/hess-2021-518-AC2, 2022

Reply on RC2

Donghoon Lee et al.

Author comment on "Unfolding the relationship between seasonal forecast skill and value in hydropower production: a global analysis" by Donghoon Lee et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-518-AC2, 2022

The paper contributes a global analysis about the value of long-term forecast for hydropower reservoirs. Specifically, the authors contrast the performance of three alternative operating schemes, basic control rules, perfect forecast-informed, and realistic forecast-informed. The latter use forecast information generated with a statistical prediction model based on four large-scale climate drivers along with local drivers (inflow and soil moisture). Results obtained for 735 hydropower reservoirs show that most dams could benefit from perfect forecasts, with these gains that strongly depend on dam characteristics; only a small number of dams however attains a performance improvement when realistic forecasts are used. The topic of the paper is absolutely timely and important, and fits nicely within HESS scope. The numerical analysis is robust and well designed, and the manuscript is clearly written. Overall, I think the paper could be a strong contribution to the ongoing debate about the relationship between forecast skill and value. Below I'm suggesting a few points to further improve the paper before accepting it for publication.

Response: We thank the reviewer for the positive comments and further critical comments that we believe have enhanced the overall quality of the manuscript.

1) the description of the dam inflow prediction model in section 3.1 is not totally clear:

1a. I did not understand is the determination of the optimal set of lead-months at line 155. Does this mean that, for each station/HP reservoir, you constructed 7 forecasts (i.e. M1 to M7) and then selected the best lead-time as the one characterized by the minimum MSE? If this interpretation is correct, how can you then run a Model Predictive Control with a 7-month prediction horizon in case the best lead-time is shorter than 7? Moreover, since forecast accuracy generally decreases with lead-time, how likely will be then selecting a lead-time longer than one month?

Response: We apologize for the misunderstanding. We constructed 7 forecast models (MP1 to MP7) for each dam, and the lead-month here refers to a lag-time of each predictor. For example, after we select four (statistically significant) predictors from lag-correlations, we choose an optimal combination of four lag-times of the predictors based on the minimum mean squared error. This applies to all 7 models independently and does not affect the prediction horizon. For clarifying this, Lines 153-156 will be changed to:

"To select the optimal lag-times of the predictors, we apply a leave-one-out cross-validation (LOOCV) scheme. Specifically, all combinations of lag-times of the predictors are cross-validated; then, the optimal set of lag-times is determined based on the minimum mean squared error (MSE)."

Regarding the forecast skill over lead-time, we found that the highest Kling-Gupta efficiency (KGE) appears in 68% of MP1, 5% of MP4, and 2% MP7 models. This is illustrated in Figure 3 and Figure 2S and explained in Lines 312-315.

1b. at line 142 the authors mention the generation of streamflow forecast for 1,200 stations, but the HP reservoirs are 735. why are you generating a higher number of forecasts wrt the reservoirs? moreover, is it correct to say you built 1,200 independent forecast models, one for each station, right?

Response: We apologize for the misunderstanding. The term "1,200 stations" refers to the prior study (Lee et al., 2018). Also, we built 7 independent MP models for each dam. To clarifying that, Lines 141-146 will be changed to:

"Our long-range inflow prediction model uses Principal Component Regression (PCR) and includes four lagged large-scale climate drivers, snowfall, and prior inflow and soil moisture conditions to predict future inflows at 735 dams. This approach is readily implemented globally and has demonstrated fair (realistic) predictive skill at 1,200 streamflow stations (Lee et al., 2018). While Lee et al. (2018) predict seasonal (3-month) streamflow averages, here we develop independent monthly prediction (MP) models for the subsequent seven calendar months. For example, forecasts issued at the end of February include monthly inflows from March (MP1) to September (MP7)."

1c. at line 138 you say that state-of-the-art physically-based forecasts fall short on lead-times up to 7 months, but actually these lead-times are covered by existing products such as ECMWF seasonal forecasts available on the Copernicus Data Store. I would thus recommend to better contextualize this point.

Response: This is a point that was also raised by reviewer #1. We agree with this comment and thus plan to modify Lines 136-141 as follows.

"Two broad alternative approaches for seasonal streamflow forecast development include physically-based models, such as GloFAS (a global-scale forecasting system; Emerton et al. (2018); Harrigan et al. (2020)), or statistical prediction models that leverage the relationship between large-scale climate drivers and local hydro-meteorological processes (Block, 2011; Gelati et al., 2014; Giuliani et al., 2019). Here, we select the second approach for two reasons. First, the prediction horizon of most currently openly available global reforecasts (a few days to 3-4 months) falls short of our preferred lead times up to seven months, needed to test the potential of realistic forecasts for a broad spectrum of reservoirs—including those characterized by slow storage dynamics. Second, re-forecasts issued by global-scale forecasting systems are only available for a relatively-short hindcast period (typically two decades; Harrigan et al. (2020)), whereas the time series of globally-available hydro-climatological data are significantly longer. It should be noted that these two statements may change in the near future as the boundaries of global-scale forecasting systems keep getting extended (see Section 5.2). For example, there already exist global re-forecasts from physically-based models with a prediction horizon of seven months and hindcast periods of about 30 years (https://hypeweb.smhi.se/explore-water/forecasts/seasonal-forecasts-global/)."

2) the labeling of dams in success/failure (section 3.3.1) based on the comparison of IPF against the average IPF raises the following question: while the definition of IPF implies

that forecast-informed operation is beneficial when IPF>0, I don't understand why a failure (i.e. IPF < mean(IPF)) implies that basic control rules and perfect forecast-informed operations generate similar amounts of hydropower (lines 251-252). According to this condition, I guess a dam can be classified as failure even if IPF > 0, right?

Response: We agree that this statement is confusing, since you rightly pointed out that a dam can be classified as 'failure' when its IPF value is positive. We thus plan to change the term from "success/failure" to "case/non-case" and to remove Lines 251-252 to avoid the confusion.

3) while I fully trust the statistical forecast model used by the author, I think the paper could benefit from some benchmarking of the resulting forecast skill against existing, physically-based forecast products. this is likely not necessary for all the models, but it could be a useful, complementary information for some representative cases, possibly selected across different climate regions.

Response: We agree with the reviewer's point on a comparison of forecast skill with physically-based forecast products. However, there are two challenges that may hinder the comparison: 1) At the majority of dams, both our statistical model and physically-based prediction methods (or products) may predict different "simulated" streamflows rather than the actual "observed" streamflows. For instance, we used streamflow data predicted by the WaterGAP model. In other words, the result of such comparison may be affected by the different characteristics of streamflow simulations (e.g., forcing data). 2) The outcome of such analysis may vary significantly depending on the composition of the subset of skilled or unskilled regions. Other minor issues include obtaining data for the exact grids of dam locations, supporting finer spatial resolution for headwater dams, and forecasting the same time period with the same lead-time.

Because of these reasons, we believe a qualitative comparison may be the best choice. To this purpose, we retrieved the performance of the Global Flood Awareness System (GloFAS), one of the most advanced physically-based streamflow forecasting systems. The figure below shows the Kling–Gupta efficiency skill score (KGESS) for GloFAS-ERA5 river discharge reanalysis against 1801 observation stations. While KGESS values are higher than the initial KGE values (Harrigan et al., 2020), the KGE scores generated in our study are comparable to or slightly higher than GloFAS scores. Even though the GloFAS's KGESS is calculated using observed streamflow, similar patterns of forecast skills can be found in our statistical forecasts, such as relatively lower forecast skills in central southern USA, southern South America, and southern East Africa, and relatively higher forecast skills in northwest North America, central South America, Europe, and South Asia.

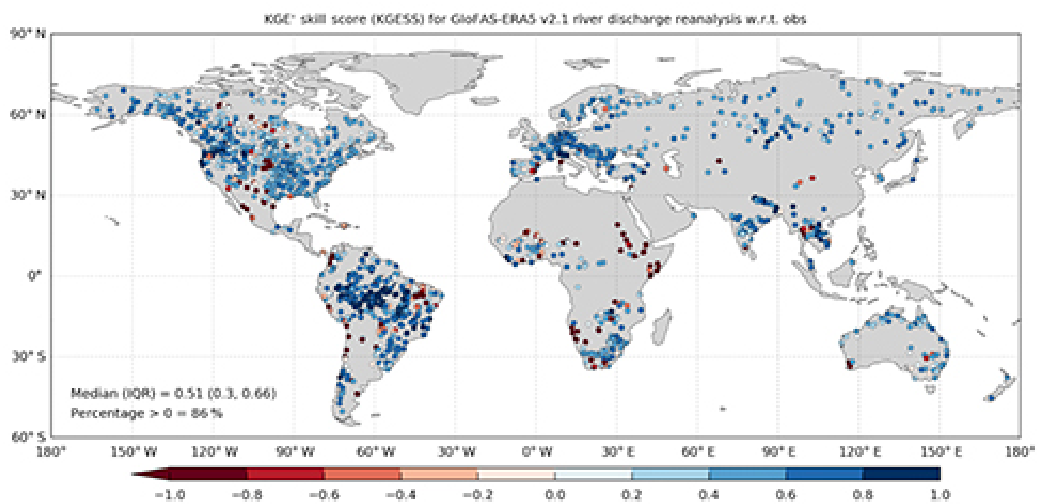

Figure 2.1. Modified Kling–Gupta efficiency skill score (KGESS) for GloFAS-ERA5 river discharge reanalysis against 1801 observation stations. Optimum value of KGESS is 1. Blue (red) dots show catchments with positive (negative) skill (Harrigan et al., 2020).

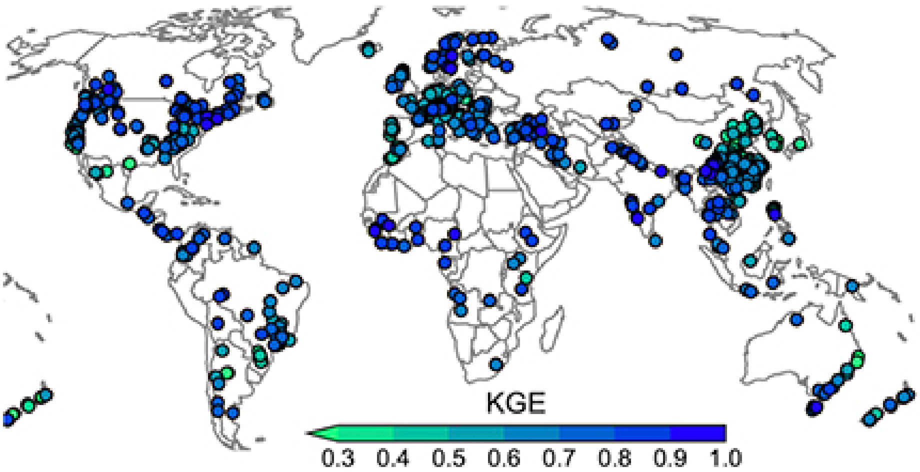

Figure 2.2. KGE scores of 1 month lead (MP1) inflow forecasts developed in our study. The original figure is Figure 3 in the manuscript.

4) the results show how the overall value of forecast information for hydropower production is (unfortunately) relatively small. Did the author consider how much is the potential influence of the experimental settings, particularly in terms of (A) informing the operation with monthly inflow forecasts and (B) assuming the reservoirs are operated to maximize total (or average) hydropower production. About (A), the work by Bertoni et al. 2021 shows how some reservoirs could benefit more from predicting the inflow peak over a given horizon, rather than the average inflow, as this information is useful in hydropower operations to avoid spilling water. About (B), I was wondering if in this context the maximization of the firm energy could benefit more than the maximization of total production as it is more related to extreme conditions.

Response: Yes, these are two points that we considered when conceptualizing the study and setting up the experiments. However, we preferred to proceed with the current setup because of a few reasons. Starting with point (A), the nature and intent of a global study require us to create a realistic setup for all reservoirs of our study site. In this regard, it is true that some reservoirs could benefit more from predicting the inflow peak (instead of the total / average inflow volume), but investigating such aspect would result in a redesign of the study, which should bank on different forecast models and different analyses of how skill and dam design specifications result in forecast value. In other words, we preferred to keep a setup that is likely to reflect what the majority of dams would benefit from. As for point (B), the rationale is similar: we opted for a setup that is likely to reflect the operational objective characterizing most reservoirs. Such choice is corroborated by the validation reported in Turner et al. (2017), where we show that maximizing total production leads to an accurate simulation of annual hydropower production. That said, we agree with the reviewer that both point (A) and (B) are relevant to our study, so we will expand our reasoning in Section 5.2.

MINOR:

- in eq. 2c, the mass balance equation includes the evaporation losses. where are these data coming from?

Response: Evaporation is calculated by multiplying the surface area of a reservoir (at each time period) by the potential evaporation. Time series of potential evaporation from 1958 *to 2000 are obtained from the Water and Global Change (WATCH) 20th century model output generated using the WaterGAP model (i.e., the same source as our time series for the inflow into each reservoir).*

References

Lee, D., Ward, P., and Block, P.: Attribution of Large-Scale Climate Patterns to Seasonal Peak-Flow and Prospects for Prediction Globally, Water Resources Research, 54, 916–938, https://doi.org/10.1002/2017WR021205, 2018.

Turner, S. W., Ng, J. Y., & Galelli, S. (2017). Examining global electricity supply vulnerability to climate change using a high-fidelity hydropower dam model. Science of the Total Environment, 590, 663-675.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., and Pappenberger, F.: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, Earth Syst. Sci. Data, 12, 2043–2060.