Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-486-AC3, 2019


Hydrology and Earth System Sciences
Discussions

# *Interactive comment on* "Inferred inflow forecast horizons guiding reservoir release decisions across the United States" *by* Sean W. D. Turner et al.

**Sean W. D. Turner et al.**

sean.turner@pnnl.gov

Received and published: 10 December 2019

**Referee Comment:** In this submitted manuscript, the authors applied a conceptual method to analyze 316 dams and reservoirs in the U.S. with respect to the roles of forecast information in driving the discharge operations. Using the proposed "horizon curves", authors specifically analyzed the relationship of forecast information and operation for four key dams, in order to test whether the proposed method could improve the modeling capability of large-scale hydrological studies with the interference of reservoirs and dams. The study was originated from the fact that water managers nowadays will rely more and more on hydrological forecasts as the information of fore-



cast range, resolution, and accuracy have been improved by various methods. However, it is still unknown how important and how influential a forecast could essentially improve the reservoirs and dams operations. One of the contributions of this study is to analyze a large number of dams in the United States based on limited reservoir operation data. In addition, a new concept of "horizon curves" has been proposed by authors. Using the proposed method, this study also tries to answer the question of "how and when do forecasts applied in the field of reservoir operations". In general, the scope of the submitted manuscript is indeed very interesting, and the main contribution novelty lies in the invention of the concept of "horizontal curve" for reservoir operations. However, the reviewer thinks there are a few key assumptions are questionable when authors develop the "horizon curve" method. Those assumptions are subject to verification and further investigation. In addition, the reviewer also finds the organization of the manuscript is confusing, and few paragraphs in the methodology section are hard to follow due to missing steps or information.

**Author Response:** Thank you very much for your time and for your constructive review. We are confident that we can address each of your concerns through clearer method description.

**Proposed changes to manuscript:** See below.

**Referee Comment:** Methodology Justification (Line 85-86) The proposed approach is based on the assumption that 1) "the future observed inflows (perfect forecast) may act as a proxy for the actual forecast available to the operator at the time of deciding how much water to release". However, in reality, reservoir operators never trust a single forecast, at least the forecast uncertainty needs to be considered when making any release decision. More importantly, most of the releases are pre-defined by the reservoir "rule curves" with limited influences from the forecasts. Regardless of the forecasts in different ranges and accuracy, reservoir operators always and have to releases a certain amount of water at a certain time following the "rule curves". Therefore, the reviewer is unclear how the forecast information based on the "horizon curves" could actually interact with the existing operating rules. The manuscript seems to omit this linkage between forecasts and the hard rules reservoir operators must follow. The rule curves are even more important in terms of mid to long-term operations, which is the same time range this study has been focused on. Different reservoirs have different rule curves, and it would vary from one reservoir to another. The proposed method of "horizon curves" seems to be a universal approach for any reservoirs. Reviewer is wondering the applicability of the "horizon curves" as each reservoir will have different settings and "rules" to follow as regulated by USACE or relevant water agencies. How does the proposed "horizon curves" could address different reservoir regulations and functions, such as hydropower reservoir vs. flood control reservoir vs. water supply reservoir vs. environmental demands?

**Author Response:** Thank you for this comment as it allows us to clarify the application of horizon curves. The horizon curves are to be implemented in a large-scale hydrological and-river routing/reservoir model for which operational rule curves are unavailable or too complex. Deriving the horizon curves is a backward process that informs simplified non-operational reservoir operations model on how to better represent reservoir releases with the assumption that foresight would help since in operations reservoir operators use them along with the rule curves. The horizon curve indicates the horizon that seems to best indicate useful information to deviate the release from climatology. We agree that we ought to show how the two-stage release function we propose aligns with rule curves and reservoirs that serve multiple functions. Let's first assume that forecasts are not used at all and that we simply want to use the storage volumes to set the release for any given week of the year (as in archetypal rule curve). Our scheme actually allows us to do this very effectively. The two-stage piecewise model provides the flexibility to represent two situations: (a) the reservoir is above the guide-curve, in which case one would expect the operator to respond by releasing water to draw the reservoir back down (with higher release for higher storage levelsâĂŤallowed for by the right-hand slope of the piecewise function); and (b) the reservoir is below the guide-curve, in which case the operator would wish to cut back the release significantly



to allow the reservoir to refill. For (b), the operator would be constrained by the fact that release cannot be negative and that some environmental flow will normally be required, so we add a constraint to our model so that the slope for (b) must be less than for (a). Assuming the operator works only to meet a guide-curve based on storage levels and time of year, our model is ideal. The breakpoint of the piecewise function on the x-axis (water availability) essentially represents the reservoir guide-curve level (since water availability is just storage volume when there is no inflow forecast). Since these rules are optimized to fit observations for each water week and for each reservoir individually, they do indeed account for the differences across the year and across dams. The model is trained on observed decisions, so is agnostic as to whether the reservoir is used for flood control, irrigation, hydropower, etc. It mimics how water is released in practice, so will in effect capture any of these purposes. We must also recognize that these rules are somewhat flexible in the sense that operators will deviate from them depending on forecast information available. Our scheme is designed to allow us to integrate forecast information in the rule-based system. The difference is that the rule is no longer a function of current storage, but becomes a function of current storage plus the inflow out to various candidate horizons, with the candidate horizon offering the best fit to observed decisions taken as the assumed operational horizon in the inferred horizon curve. We agree that operators are more likely in practice to use a forecast ensemble than a deterministic forecast. Given the large scale of our study and the intended application (ultimately a CONUS-scale hydrological and river-routing model) we require a simple proxy for the forecast information available to the operator. Our results demonstrate that the actual, future inflow serves this purpose well in many cases.

**Proposed changes to manuscript:** We will amend the method section to explain how our chosen model is compatible with reservoir guide-curves used in real-world operations for a variety of purposes.

**Referee Comment:** The steps of horizon curve method (Line 80-120) The reviewer cannot fully understand the whole process of deriving the horizon curve. For example, in line 98, the authors mentioned the "release-availability" function for the first time in the manuscript and then briefly explained the definition of "availability a". How do the authors define "release-availability" here? How did authors construct the functions of "release- availability"? The possible releases must be from the existing operating rules to prevent overtopping and dead-pool of reservoir storage. Where do authors obtain such information in a national scale? This is a term again not commonly used by water managers, and more explanations would be needed. The authors also wrote, "The inferred policy function fitted to these data is a piecewise linear model with a single breakpoint" in line 99 and the reviewer is wondering what does the "policy function" here refer to? And what do "these data" referred to as? In general, the reviewer thinks this section is hard to follow given lots of non-common terms were used. Those wordings may make sense to authors themselves, however, it is not apparent to water managers and operators. Please re-check some of the literature and especially reservoir operation reports to further explain how the proposed method is constructed in detail. A flowchart or additional figure may be added to explain the steps of creating this horizon curves.

**Author Response:** Our model is indeed quite complex. Figure 1 was designed to help the reader in this regard, but on reflection we agree that the explanation would be enhanced with a flow chart outlining the process and data used at each stage. The release-availability function is the key to the whole approach. This is simply a relationship (derived for each water week and defined by a broken linear model with two lines and two slopes) that specifies the water release as a function of water availability. These are constructed using observed records of release, storage, and inflowâĂŤso the releases are from existing operations designed to prevent overtopping and dead-pool of reservoir storage. We obtain these records from five sources: US Army Corps of Engineers, US Bureau of Reclamation, US Geological Survey, California Data Exchange, and Texas Water Development Board. These are listed in the experimental setup of the manuscript. We use "policy function" and "release-availability" function



interchangeably, which we agree is confusing for readers.

**Proposed changes to manuscript:** We will define the release-availability function more clearly and will add a flow diagram to guide the reader through process of deriving the horizon curve.

**Referee Comment:** The use of Random Forest Classifier (Line 160 - 170): There are few nested issues about the description on the use of Random Forest Classifier, experiment setting, and how will these experiment settings lead to a conclusion related to the forecast information and reservoir operation. First, authors should point out what are the "features" and "target" when using Random Forest Classifier. The authors mentioned there are 26 features without a tabular form to let readers know what those are. In addition, the "target" used in RF is still unclear. Did the authors intend to figure out which feature has the most important influences on release decisions? Or did the authors intend to classify reservoirs according to their correlation between discharge and inflow fore- casts? And how was this realized in RF? The description here reads very short and is not comprehensive. Reviewer is confused about what has been classified based on what inputs, as well as how this experiment setting would lead to a certain conclusion. At least a few additional paragraphs would be necessary to explain the experiments here.

**Author Response:** Thank you. We did indeed define the target, but we rather sloppily referred to it as the "response variable" in the method. From line 164: "The response variable is a Boolean of whether there is evidence for significant forecast contribution or not." It's a simple TRUE/FALSE for whether there is evidence for forecast use in the horizon curve of each dam. The point of the RF analysis is thus to understand if there are features of dams (storage capacity, etc) that would be associated with detected forecast use. We expect that with this clarified this part of the study will be much clearer and perhaps doesn't require the additional few paragraphs suggested.

**Proposed changes to manuscript:** We will list all 26 input features of the RF classification and adjust the wording for consistency throughout method and results.

**Referee Comment:** data segmentation (Line 160 - 170): Since the methodology used here is Random Forests as one of the machine learning tools, the reviewer is wondering whether there is an overfitting issue? Traditionally, the data should be split into training, validation and test periods to verify there is no overfitting. However, authors only did a training and a test without validation. It is likely the model is overfitted and more experiments on different folds are necessary to justify the proper use of random forests.

**Author Response:** The traditional machine learning methodology uses training data to train a model, uses validation data to tune model hyperparameters, and uses testing data to evaluate model performance. We agree that if a sole testing set is used for both hyperparameter tuning and model evaluation, the model may be biased and prone to over-fitting. But in our case, the only two hyperparameters tuned (number of trees and maximum tree depth) are not adjusted to achieve best testing score, but instead determined in reflection of our small dataset: maximum tree depth is limited to 3 levels to reduce model complexity, and we choose a considerably large number of trees of 1000 to reduces the model variance and hence reduce overfitting (the larger the number of trees, the lower the ensemble model variance). Additionally, we repeated the experiments 200 times with different splits, each time train the model from scratch on training data and evaluated the model performance only on unseen testing data. We are therefore confident the model is not overfitted.

**Proposed changes to manuscript:** We will add the additional details of the Random Forest experiment design into the draft.

**Referee Comment:** Gini index Line 305: Can the authors define what is the “Gini impurity of the tree”? Some examples and references using this index would be helpful.

**Author Response:** Thanks for the request for clarification. Gini impurity, like entropy, describes the heterogeneous state of a system and is formally calculated as $\sum_{i=0,1} P_i * (1 - P_i)$ in a binary classification tree node, where $P_i$ is the fraction of samples that



belongs to class $i$ in this particular node. The feature importance score is calculated as the percentage of Gini impurity decrease because of the particular feature, averaged across the forest. The sum of all feature importance scores equal to 1.

**Proposed changes to manuscript:** We will explain Gini impurity and how feature importance scores are calculated. We will also add the equation and reference to explain the Gini impurity concept.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-486, 2019.