Technical report on Explainability and Communication approach

Water Supply Forecast Rodeo: Final Stage

By @iamo_team

Abstract

This report describes the content, data, and methodologies, we used to communicate seasonal streamflow forecasts. For each issue date, the forecast summaries present predictions of seasonal streamflow volumes putting them in historical context, alongside current conditions for key hydrological variables like basin snowpack, accumulated precipitation, and antecedent river flow. These summaries offer insights into the evolving prediction power of the used variables over the target water year, which is important for understanding forecasted streamflow volumes. We also address variations in the predictive power of these variables along all issue dates, accounting for factors like basin characteristics and issue date timing by including information on their relevance derived from the forecast model using a model-agnostic method.

Technical Approach

- **Forecast and uncertainty communication**

Our forecast solution involves training a PLS model for each issue date per each river, using a predefined range of PLS components and evaluated using Leave-One-Out Cross-Validation (LOOCV). The suggested forecast summaries communicate predictions of seasonal streamflow volumes, preceded by summaries of current conditions for key hydrological variables used as predictors in their respective forecast models.

The seasonal terrestrial storage accumulated during winter time in the western basins of the US is considered the main source of predictability for river discharge during the summer season. Therefore, primary predictors in the forecast models comprise of Snow Water Equivalent (SWE) and accumulated precipitation from the Snow Telemetry (SNOTEL) network. Another significant predictor we use is the antecedent (naturalized) river flow. Therefore, each summary begins with information on the current conditions of these variables with the help of graphs depicting their propagation since beginning of target water year up to forecast issue date. In addition the summary includes a variable importance plot derived from the underlying model specific for river and issue date.

Following the overview of current hydrological conditions, the summaries present forecast model outputs as a one-line table containing predictions at the 0.10, 0.50, and 0.90 quantiles. To provide users with context regarding the median prediction (0.50 quantile) for the current season, we supplement the table with the median seasonal streamflow volume for the 1990-2022 period. Additionally, we include the seasonal streamflow volume for the previous water year in the table, as users may find it helpful to compare recent hydrological conditions.

The information is displayed in this deductive manner to offer users a transparent view of the forecast process. By starting with summaries of current conditions for key hydrological variables, users can grasp the immediate context in which the forecasts are made. Understanding the current state of variables such as SWE, accumulated precipitation, and antecedent river flow provides crucial insights into the factors influencing streamflow in upcoming months. The accompanying graphs depicting the propagation of these variables since the beginning of the target water year offer a visual representation of their deviations from long-term medians, aiding in the interpretation of their current status.

- **Explainability metrics and communication**

The primary content of each summary, in addition to quantile predictions, includes the current conditions of key hydrological variables utilized by forecast models. These variables encompass SWE accumulated precipitation, and antecedent flow. Such information offers crucial context for understanding the forecasted streamflow volumes. Figure 1 provides as a reference a snapshot of 15-May-2023 forecast summary for the Owyhee River.

The basin snowpack summary depicts current level of SWE and how it has been progressing since the start of the current water year. To provide a historical reference for comparison, the graph also contains information on median SWE as well as SWE for the preceding WY. The length of training data differs across rivers, ranging from 33 to 38 years of observations, because of different period of availability of SNOTEL data. To allow for comparability across all rivers we maintain a single reference period for calculating median SWE, from 1990 trough 2022. In addition to the median SWE, the corresponding graph also depicts the spread between the maximum and minimum values, as well as the interquartile range (IQR) of SWE over the reference period.

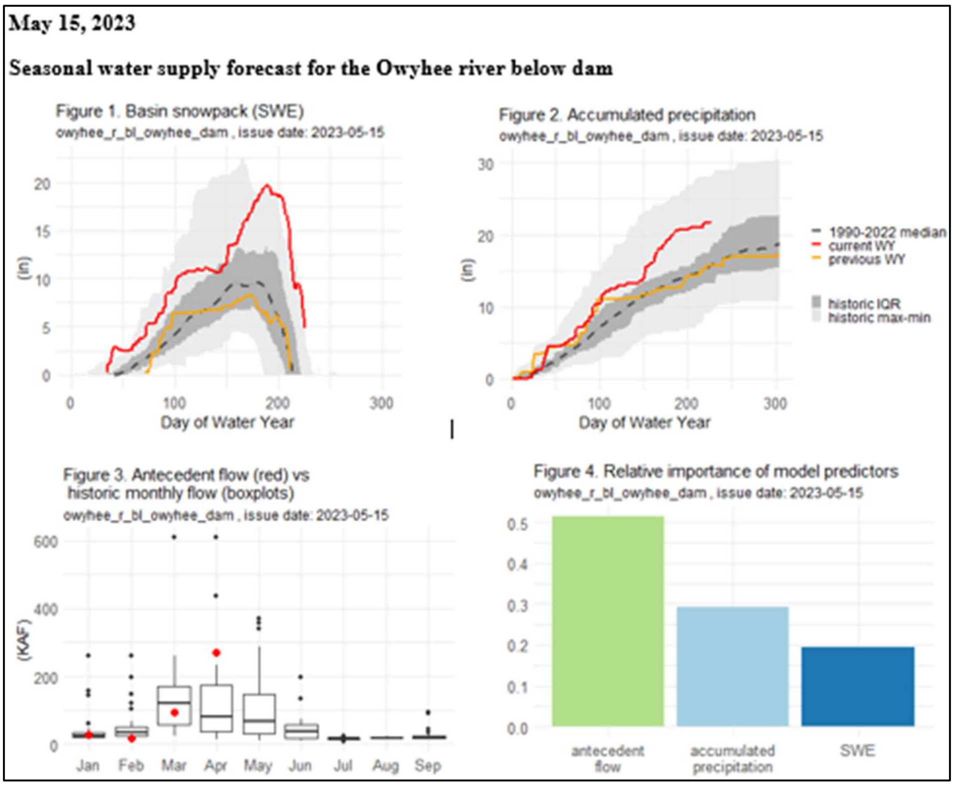

Figure 1. Example of forecast summary visuals for 15-May-2023 issue date for Owyhee river

Our forecast solution assimilates SWE records from up to four selected SNOTEL stations per basin, as well as basin-averaged SWE estimates from the SWANN dataset. In large basins with complex terrain and a wide range of elevations, estimates from different SNOTEL stations may show varying trends. This presents a challenge, as we need to present a single graph that accurately represents basin-wide snowpack conditions. Ideally, we would use basin-averaged SWANN estimates for this assessment. However, the computational burden and the lack of a readily available package for the automatic download of SWANN data have prevented us from using it as the primary source. Therefore, to represent basin snowpack conditions, in each basin the solution uses a SNOTEL station with SWE observations that have the highest correlation with seasonal streamflow volume.

The summary on accumulated precipitation provides information in the same manner as for SWE: offering insight into how precipitation has been accumulating during the target water year up to the target issue date, with contextual information on historic max-min and IQR spreads. As with SWE estimates, we use precipitation records from the SNOTEL stations that have the highest correlation with historical streamflow volume.

Although it has inferior predictive power at early issue dates, antecedent streamflow becomes a major predictor in most rivers by late issue dates due to the strong autocorrelation of monthly flow volumes from late spring through summer months. Therefore, we supplement the forecast summary with visual information on monthly streamflow volumes prior to the issue date. This includes a plot depicting the levels of observed streamflow volumes before the issue date, with a hydrograph of the monthly streamflow at the background. The hydrograph provides information on streamflow median, interquartile range (IQR), and max-min values during the reference period (1990-2022).

Our forecast solution includes using climate oscillations indices, beyond the noted variables that proxy current hydrological conditions, as precursors of future hydroclimatic conditions. These include indices of Southern Oscillation Index (SOI), Pacific Decadal Oscillation (PDO),and Pacific-North American pattern (PNA). Inclusion of climate indices proved useful for majority of the rivers particularly at the earliest forecast issue dates. Visually placing the current state of climate indices in a historical context is problematic because each river's streamflow association is distinct, and can exhibit inverse relationships across the western US in both space and time. Instead of present it as a graph, we therefore suggest to include in the forecast summaries for early issue dates, an information about current state of relevant climate oscillations and how they are associated with seasonal streamflow of considered river. However, the submitted forecast summaries omit such descriptions because the forecast issue dates fall on later dates, when basin-specific climate indices have little to no predictive power.

All noted types of predictors—climate indices, SWE, accumulated precipitation, and antecedent streamflow—have varying degrees of predictive power depending on the river, most likely due to basin climatic characteristics, latitudinal, and altitudinal locations. Furthermore, their predictive importance changes over time within a single basin. As it was mentioned, climate indices may be useful at early issue dates, while SWE and accumulated precipitation usually show a strongest relationship with anticipated streamflow in spring. During the summer issue dates, in the most study rivers, expected seasonal streamflow volume can be robustly predicted by using only antecedent streamflow over preceding months. Without understanding such basin and subseason-specific peculiarities, information on current state of the noted variables may be confusing, especially when they show divergent trends. E.g. SWE lower than the long-term median in combination with above-normal antecedent streamflow may indicate two very different expectations for seasonal streamflow volume: which depends on whether this combination is observed during early or late issue dates.

In this regard, we supplement the forecast summary for each river with additional information that describes which of the considered variables are more relevant for anticipated seasonal streamflow given a current issue date. To achieve this, we derive information on variable importance from the underlying forecast model for each river at each issue date. While there is a straightforward approach for extracting variable importance from Partial Least Squares Regression, which involves extracting coefficients for each variable resulting in PLS components, our solution encounters distortion issues due to highly correlated variables (e.g., SWE or precipitation from up to four SNOTEL stations in each basin). Instead, we employ a model-agnostic method based on a simple *feature importance ranking measure* after Goldstein et al. (2015), which quantifies the effects of each feature by using partial dependence plots. This method assigns relative importance scores to each variable used by the model, which we then summarize into four classes according to predictor type: climate indices, SWE, accumulated precipitation, and antecedent streamflow.

References:

Goldstein, Alex, Adam Kapelner, Justin Bleich, and Emil Pitkin. 2015. "Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation." Journal of Computational and Graphical Statistics 24 (1): 44–65. https://doi.org/10.1080/10618600.2014.907095.