Water Supply Forecast Explainability Report

Abstract

The explainability reports are provided as a supplement to the seasonal naturalized streamflow quantile prediction model, to offer contextual data and data visualizations that help users of the model understand what data is most influential for a stream site and how the model weighs that data in obtaining the predicted value. The architecture underlying the model is an ensemble of up to 12 different models. The explainability report includes visualizations that offer insights from individual components of the model ensemble and visualizations that show how the contributions of the ensemble come together. We benefited from existing work that provides explanations of gradient boosted models using Shapley values and further customized these explanations to suit the ensemble of quantile regressors. The authors' hope is that consumers of the explainability report will obtain an understanding of both how and why the model produced a set of predicted quantile values on a particular date and be able to reason about how changes in conditions for a stream site will impact predictions in the future.

Brief Summary of the Model

The architecture of the model consists of three separate ensembles of 4 gradient boosted quantile regressors, one for each quantile 10th, 50th, 90th. For each quantile there is both a monthly model (if monthly naturalized flow is available) and a seasonal model, each of which consist of a LightGBM implementation and a Catboost implementation. The ensemble model architecture was chosen for accuracy reasons; it is the model that resulted in the lowest mean quantile error during cross-validation. Ensembling models adds complexity to model explainability, so great lengths have gone into visualizing how each component of the ensemble model uses data and how those components are combined to arrive at the final model prediction.

Brief Summary of each Figure in the Forecast Explainability Report

Each of the plots in the Forecast Explainability Report are unique to the predictions for a specific stream site on a specific issue date. This document will focus on 4 example Forecast Explainability reports: owyhee_r_bl_owyhee_dam (owyhee) 2023-03-15 & 2023-05-15 (owyhee-3-15/owyhee-5-15), pueblo_reservoir_inflow (pueblo) 2023-03-15 & 2023-05-15 (pueblo-3-15/pueblo-5-15).

Fig 1 is a stacked bar graph with the y-axis representing the predicted naturalized streamflow volume (KAF) for the current issue date. The x-axis represents each quantile (10, 50, 90). For each predicted quantile there are 4 sub-models in the ensemble that contribute to the prediction and the stacked bars are proportional to the sub-models contribution for that prediction. This visualization is meant to concisely show the predicted quantiles for the current issue date and how each of the sub-models in the ensemble contributed to the overall prediction. For stream sites without monthly data there will be only 2 sub-models for each quantile.

Fig 2 plots all of the predictions made in a streamflow season up to the current issue date, with a solid line for each quantile. There is also a dashed line for each quantile which represents the historical quantile value using data from previous years streamflow volume. This visualization is meant to show how the predictions have evolved over the streamflow season and how the predictions compare to streamflow volumes observed in the past for the stream site.

Fig 3 plots the observations of the dynamic features (features with values that change within the streamflow season) for each prediction up to the current issue date. Fig 3 is plotted directly under Fig 2 to compare how the quantile predictions change as the feature values change. The values plotted are z-scores, representing the standard deviation of the observed value from historical values captured at the same site and at the same time of year.

Fig 4 plots the current predictions for each streamflow month produced by the models that predict monthly streamflow and include dashed lines that show the historical monthly quantiles. This plot gives a picture of how streamflow volume progresses within a season for a given stream site. For forecasts later in the streamflow season, when the issue date is beyond months within the streamflow season, the observed naturalized monthly volume is used by the model for each quantile. In the plot this manifests as a collapse of all of the quantiles onto a single point which shows that there is no longer uncertainty for the predicted volume in that month. This figure will not be available for stream sites that do not have monthly naturalized flow observations.

Fig 5 plots a heat map of the change in the z-score features from the previous prediction. This plot works in combination with Fig 7-9 which shows how the predictions changed in the monthly models from the previous issue date. Combined these plots help to explain how changes in the features from week to week impact the model predictions. Fig 6 shows the value of the z-score features from the current prediction.

Fig 7-9 highlight the monthly models and make use of an explainability tool for decision tree based algorithms called Shapley Values.[1] There is a figure for each quantile. On the left side of the figure is a plot that shows the difference in the Shapley values between the prediction for the current issue date and the previous issue date. There is a point plotted for the Shapley value change corresponding to the prediction for each month within the streamflow season and for each dynamic feature category. On the right side of the figure is the predicted monthly naturalized streamflow volume for each month in the streamflow season. The bar represents the prediction for the current issue date and the circle represents the prediction for the previous issue date. These figures will not be available for stream sites that do not have monthly naturalized flow observations but could be replaced with similar plots for the seasonal model.

---

[1] https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html

Fig 10-12 highlight the seasonal (yearly) streamflow model subset and also make use of Shapley values. There is a figure for each quantile. In these plots, the Shapley values for the current issue date prediction are displayed for each of the feature categories and a directional bar shows how the value of that feature contributed to the predicted log volume. The direction of the Shapley values and their relative magnitude explain how the observed value for that feature group contributed to the predicted volume. Because Shapley values are additive, this plot encourages the reader to imagine how all of the contributions of the features produce the prediction. On the bottom of the plot is a value for E[f(x)] which represents the expected value of seasonal log volume for all observations of streamflow in the training data, this is an average across all stream sites. On the top of the plot is a value for f(x) which represents the predicted value of seasonal log volume for the stream site featured in the explainability report. For stream sites that have lower streamflow volume than the average across all stream sites in the training data it is common to see a negative Shapley value for the static site features, to represent that the model expects lower streamflow volume for that particular site. Static features do not change during the streamflow season, like elevation of the stream site and area of the drainage basin. By ignoring the Shapley value for the static features the reader can get a picture of how the observed dynamic features of the stream site impact the streamflow prediction from the baseline (historical average volume).

Fig 13 focuses on displaying information for a particularly influential feature, the Snotel snow water equivalent (SWE). The plot shows the geographical boundaries of the drainage basin for the stream site, with the stream site marked by a yellow X. It also includes the locations of the Snotel stations that are averaged into the combined_swe feature used by the model. The color of each location represents the z-score for SWE at that particular site. For additional context the relevant state boundaries of the Snotel stations and drainage basin are also plotted. A similar plot could be made to highlight the locations of PDSI measures, weather stations, or any geographically localized measurement.

Technical Approach

Forecast and uncertainty communication

One of the objectives of this challenge was to communicate the uncertainty in the predictions of naturalized streamflow. The uncertainty takes the form of predicting a value for the 10th quantile, a value which we expect to exceed the actual streamflow 10% of the time, and the 90th quantile, a value which we expect to exceed the actual streamflow 90% of the time.

The model we used to predict these values was an ensemble of decision tree based quantile regressors. To demonstrate the contribution from each model type to the ensemble prediction we plot a stacked bar graph (Fig 1) for each quantile which shows the overall quantile prediction and the relative contribution for each model. For model’s early in the streamflow season we expect the models based on yearly streamflow data to have a greater contribution and towards the end of the season the monthly model contribution increases as naturalized streamflow volume from months within the streamflow season is observed. In both the owyhee and pueblo reports we can see the proportion of the monthly model is greater for predictions on 5-15 than on 3-15, especially in the 10th quantile.

In Fig 2 we plot all of the predictions made in the current streamflow season and include dashed lines that show the quantiles based on historical data. This plot allows the viewer to see how the uncertainty in the prediction has changed through the streamflow season. One dramatic example of this is the large reduction in uncertainty on 2023-05-01 in the owyhee-5-15 report. Fig 4 shows that historically the highest median streamflow month for owyhee is April. On 5-1 the naturalized streamflow for April is observed so the quantiles collapse onto a single point and this greatly reduces the uncertainty of the predictions (Fig2/Fig4). While a similar process occurs in the pueblo-5-15 report, Fig 4 shows that April is the month of lowest median streamflow and therefore observing the monthly value for April does not reduce as much uncertainty for pueblo.

Because each of the quantiles is modeled with a separate quantile regressor we can expose the unique impact of feature categories on each quantile. In pueblo-3-15, we observe that the SWE measurement z-score is near 0 (Fig 6), meaning that SWE is similar to the mean value historically observed for that stream site. In Fig 10, we see that having a SWE z-score of 0 leads to an increase in the prediction for the 10th quantile. However, the same SWE z-score value has a much smaller impact on the predictions for the 50th and 90th quantile. This suggests that a SWE measurement that is similar to the historical mean for pueblo in mid-March raises the floor (lower bound) for the seasonal streamflow volume prediction but does not raise the expected median or upper bound as much.

Explainability metrics and communication

While we constructed our features and designed the models with the main goal of obtaining the most accurate prediction of naturalized streamflow possible, the decisions also facilitated features amenable to explainable predictions. There are two main types of features used in the models, static features and dynamic features. Static features do not change during the streamflow season, like elevation of the stream site and area of the drainage basin. The dynamic features are features with values that change throughout the streamflow season.

The most important dynamic features in the model are SWE (snow water equivalent), streamflow (stream gage measurement - not naturalized), precipitation (recorded from weather stations), max daily temperature (recorded from weather stations and Snotel sites), and PDSI measurements (Palmer Drought Severity Index). To obtain the z-score features we calculate the mean and standard deviation of the raw values with respect to a measurement location and the day of the year. The mean and std are used to obtain a z-score. The z-score represents the number of standard deviations from the mean for that measurement during that time of the year. A rolling average of the z-score values is taken for a period ranging from 10 to 180 days. The z-score allows for easy comparison of the relative value of dynamic measures, both between stream sites and between stream seasons for the same site, which makes it an especially explainable representation for the features that generalizes across stream sites and seasons.

The z-score features are highlighted in several of the visualizations in the explainability report. In Fig3 of the pueblo-3-15 report we see a large increase in the z-score for precipitation from 3-8 to 3-15 and we see a corresponding increase in the predicted quantiles (Fig2). Fig7-9 show that the feature with the greatest increase in Shapley value from 3-8 to 3-15 was precipitation, further confirming that the observed precipitation drove the increase in predicted streamflow volume.

In pueblo-5-15 we see that precipitation continued to be above average in March and early April but settled to normal levels in early May (Fig 3). This caused the predicted values of streamflow volume to drop in late April. However, between 3-8 and 3-15 we saw an increase in the streamflow gage measurement and accumulated water measurement (Fig3/Fig5). Fig7-9 further confirms that accumulated water and streamflow observations were the biggest driver of the increase in predicted volume between 3-8 and 3-15. A more complete story illuminated by Fig2/3 in pueblo-5-15 is that the precipitation observation is a leading indicator of streamflow volume, but the effect is ephemeral if observed early in the streamflow season (before the streamflow season window). Streamflow gage measurement and accumulated water measurements lag precipitation measurements and therefore may only start to impact the model predictions several weeks after the preceding precipitation occurred.

While numbers, line, and bar graphs can paint a detailed picture of streamflow volume dynamics; it is important to remind ourselves that stream sites are physical locations with importance to their surrounding geographies. Fig13 focuses on geography, outlining the boundary of the streamflow drainage basin and the related states. The points represent the physical location of the most correlated Snotel stations. The color of the point represents the z-score for the SWE measurement on the issue date at that particular station. SWE was chosen because it tends to be the most important feature for the majority of stream sites, however, this figure could be populated with the location of weather stations, drought measurements, etc.

In most cases, the chosen stations are within or very near the drainage basin, like pueblo-3-15 and 5-15. In some cases the correlated stations appear further from the drainage basin, like owyhee-3-15 and 5-15. One interpretation is that there is a deficit of Snotel stations near the owyhee basin. This signals to water resource planners that more stations near the drainage basin could improve streamflow predictions. Another interpretation is that geographically distant SWE measurements that correlate with a site’s streamflow volume share seasonal weather patterns, elevation, and surrounding geography and thus can be very useful to streamflow volume prediction.

Conclusion

Better seasonal streamflow volume predictions and an understanding of the dynamics that create seasonal streamflow volume allows decision makers to most effectively allocate scarce water resources. The visualizations presented in the explainability report not only help to explain how and why the model makes the predictions, they also help to illuminate areas where the model could be improved through better data collection and feature engineering.