01/30/2026

Jereme Gaeta (CDFW) Response to Reviewer Request

Feature Importance

Given that each model run is comprised of 5,000 regression trees and we are using 30 model runs, quantifying the relative contribution of each feature across these 150,000 regression trees is complicated. However, “variable importance” (hereafter, “feature importance”) is a useful approach to synthesize each feature’s importance integrated across the entire study period into a single metric. Feature importance is calculated by assessing the classification error of each tree with and without the feature of interest. These values are then averaged across all trees and normalized. Generally speaking, a feature’s importance is a quantification of a model’s reliance on a given feature to assign the correct classification.

Our feature importance analysis indicates that features related to water temperature, the amount of winter-run Chinook salmon in the system, and the brood day-of-year (DOY) are the most important predictors of salvage (Figure 1). Furthermore, we see that the monthly average of water temperature where the Sacramento River enters the Delta (i.e., Sacramento River near Sherwood Harbor) approximately three weeks prior is always the most important feature across all 30 models. However, we must note that the low importance values of all features of high management relevance (denoted by blue in Figure 1) does not mean these features are irrelevant but, rather, that other variables, such as seasonality as encompassed by water temperature and brood DOY, are *more* important when integrated across the entire year. This could simply be the fact that features of high management relevance, such as exports or San Joaquin River flow, may be important to the routing of winter-run Chinook salmon or important during very specific times of the year or circumstances, but the influence of these features are far less than the features that are associated with winter-run Chinook salmon being present in the Delta (e.g., the time of year as indicated by water temperature and brood DOY) or with strong or weak year-classes (e.g., the pass estimate and catch). In other words, if there are no winter-run Chinook salmon in the Delta during a particular season (i.e., late spring through fall), no amount of exports, whether high or low, will have an effect on whether individuals are detected in salvage. Understanding the effect of model features on a daily basis requires an alternative metric known as shapely additive explanations.

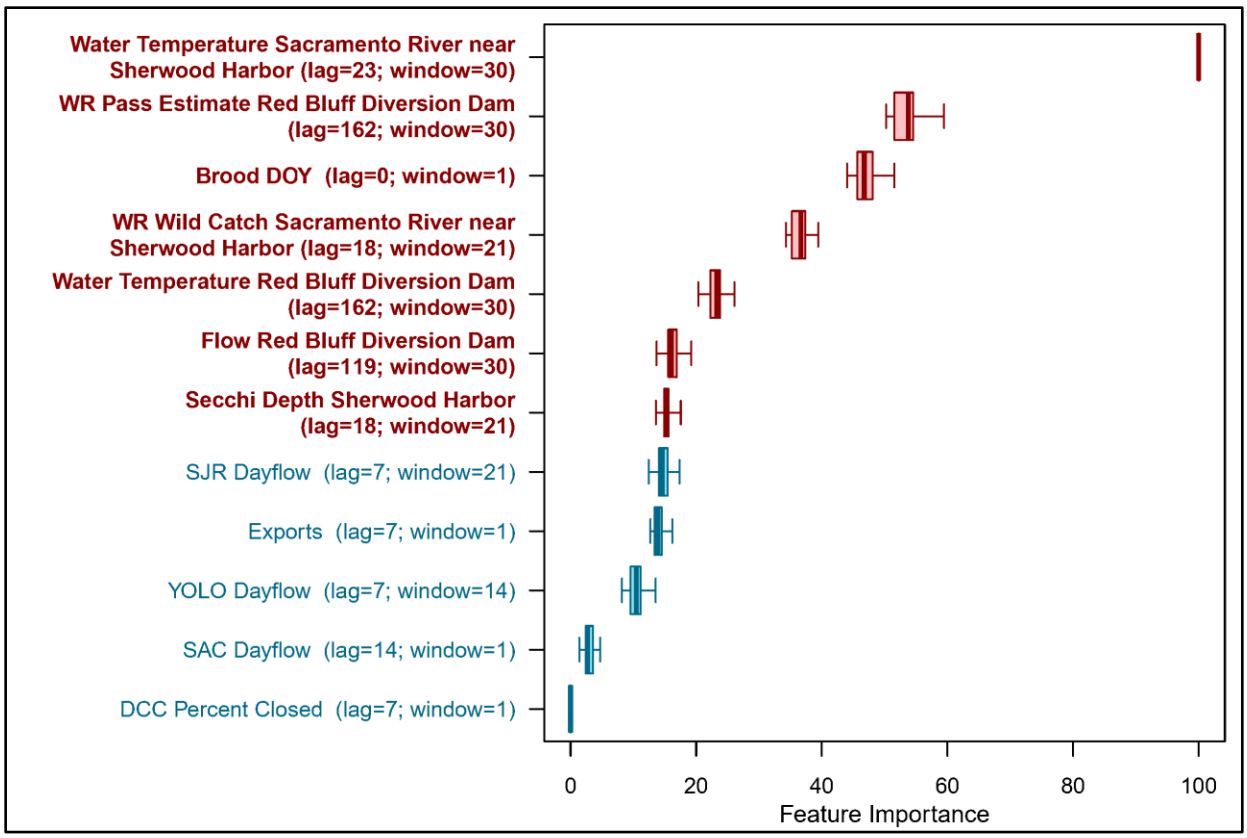

Figure 1. Feature importance of the final feature set given thirty model runs with unique starting seeds. Feature selection was based on variable importance values ⩾ 15% of the maximum feature importance in a previous, full feature analysis (denoted by red) or were identified by the stakeholder sub-group and water operators as features of high management relevance (denoted by blue).

Shapely Additive Explanations

SHapley **A**dditive ex**P**lanations (SHAP) are a common approach to evaluate the contribution of each feature to machine learning model daily outcomes (Lipovetsky and Conklin 2001). In short, each outcome (i.e., each model prediction) is estimated by the additive contributions of each feature's observation. A SHAP value illustrates the relative contributions of each feature observation to the outcome with a large *positive* SHAP value indicating the feature observation contributed strongly toward an increase in the probability of the outcome (i.e., absence, low presence, or high presence), a large *negative* SHAP value indicating the feature observation contributed strongly toward a decrease in the probability of the outcome, and a SHAP value approaching zero indicating the feature observation contributed weakly or did not contribute toward the probability of the outcome.

The distinction between the classification probability and the SHAP is worth noting. The classification probability (i.e., the probability of absence, low presence, or high presence) indicates ***what*** the model predicted as the outcome while the SHAP indicates ***how*** the model features contributed to the outcome. So, while the sum of SHAP values across all features is correlated to the classification probability, they are not interchangeable.

We visualized each feature's contribution (as indicated by the SHAP value on a given day) to the classification of "absence" in a force plot during the winter-run Chinook salmon season in 2020 (Figure 2). In this example, we see that two of the most important predictors based on feature importance (Figure 1), the Red Bluff Diversion Dam pass estimate and brood day-of-year, contributed weakly to the classification of absence on this particular day whereas one of the lower importance variables based on feature importance, exports, is contributing strongly to the overall classification of absence. This highlights that limiting our interpretation of the value of a feature to the feature importance plot alone (e.g., Figure 1) may not paint a complete picture of the model dynamics.

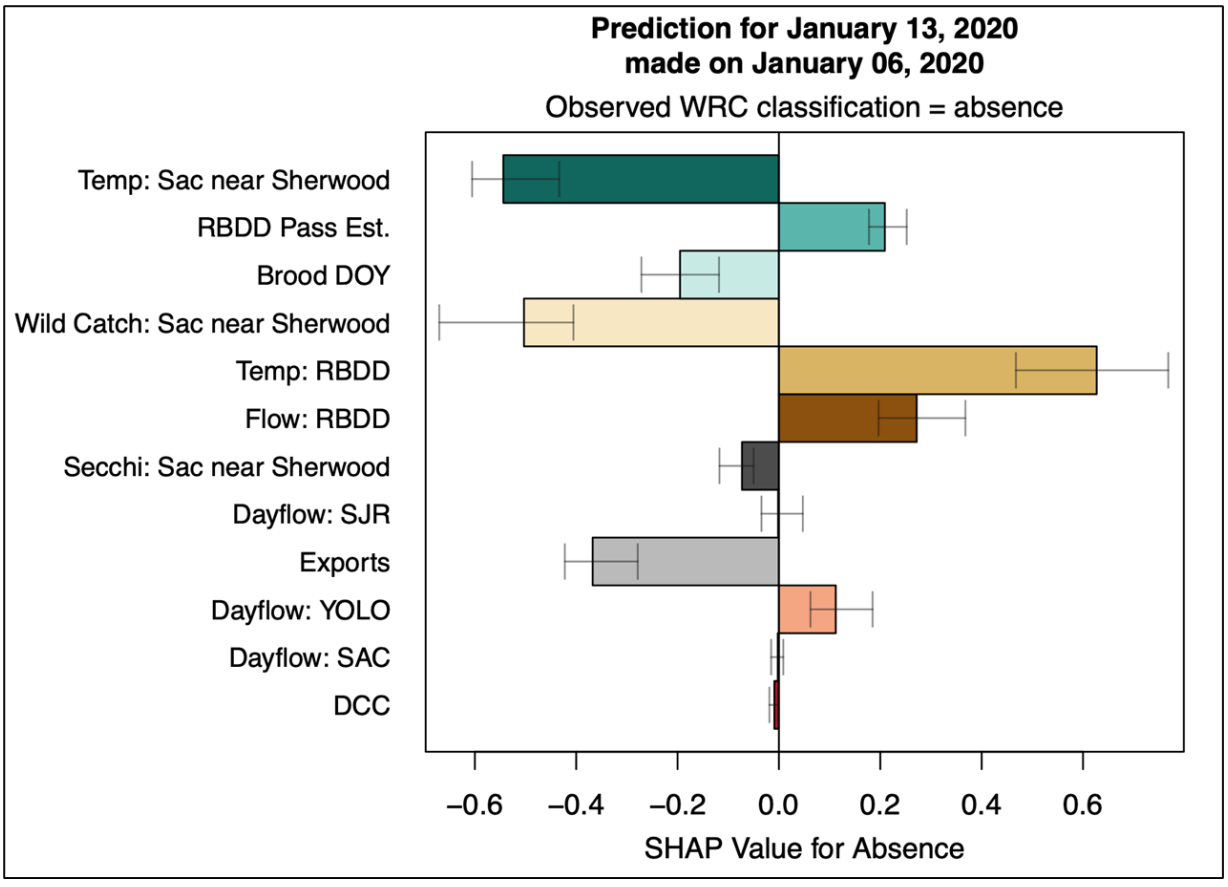

Figure 2. SHAP feature contribution to the probability of a classification of absence across 30 model runs (bar = median; error bars = range) for predictions made on January 6, 2020.

01/30/2026

Features are sorted by overall feature importance. Negative SHAP values indicate the feature is contributing toward a reduction in the probability of Absence; positive SHAP values indicate the feature is contributing toward an increase in the probability of Absence.