12/11/2025

Jereme Gaeta (CDFW) response to panel questions

Questions for the authors of the Winter Run Chinook Machine Learning Model from the review panel (Rich Zabel, Anna Sturock, and Nancy Monsen)

Is there more documentation available? We have some specific questions about which predictive parameters were included in the full model, how lags and windows were determined, and which model selection routine was used to eliminate parameters, among other questions

---

How did the authors choose the particular approach to use (i.e., the XGB-DART approach)?

At the time of model development, XGB-DART was gaining prominence, winning multiple AI modeling competitions, bringing the method to our attention. More importantly, the method had several qualities that made it ideally suited to our analytical goals. First, the method is non-parametric, non-linear, and can handle complex interactions. Second, the method generates SHAP values which allow the end user to determine what the model is "thinking". That is, you can look "under the hood" and see how each predictor contributed to the outcome on any given day (i.e., are the exports contributing toward an increase in the probability of salvage?). Third, the method allows for multiple classifications (i.e., absence, low presence, or high presence in salvage).

Can you please clarify this statement: "We used an 80%/20% training/testing data split and divided the dataset into slices, resulting in five, 1,225-day slices comprised of 980 training days followed by 245 testing days." What are the day slices composed of? Did you split the data into training/testing for each model run?

Each model run used the same testing/training split. Time slices are a common approach for timeseries modeling in machine learning frameworks. The data are broken into equal slices across the timeseries and each slice is then broken into testing and training data. We selected five time slices because this not only improved model fit, but each testing component of each time slice spanned the temporal period in which winter-run Chinook salmon are present in the Delta. Below are two figures from an older version of the model that used four time slices:

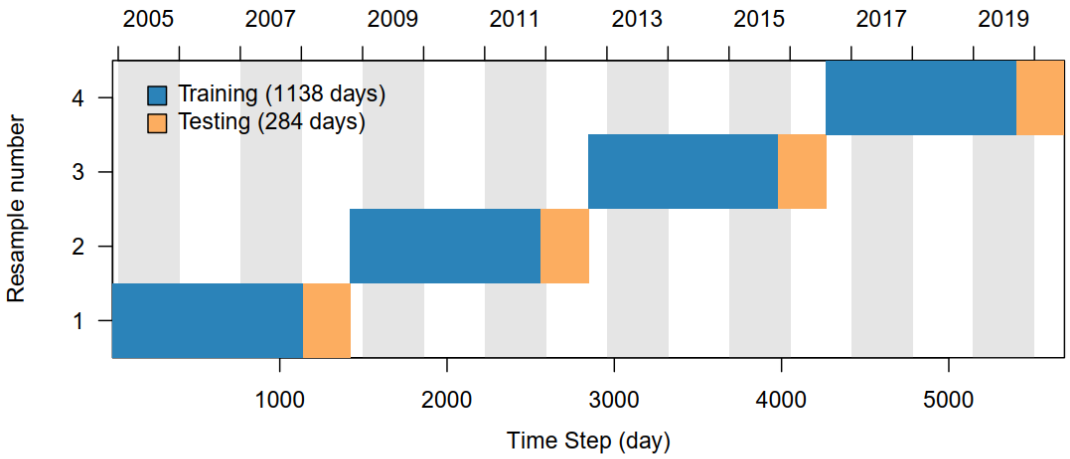

Figure 2: Time slice structure given four slices of 1,422 days each with an 80%/20% training/testing data split.

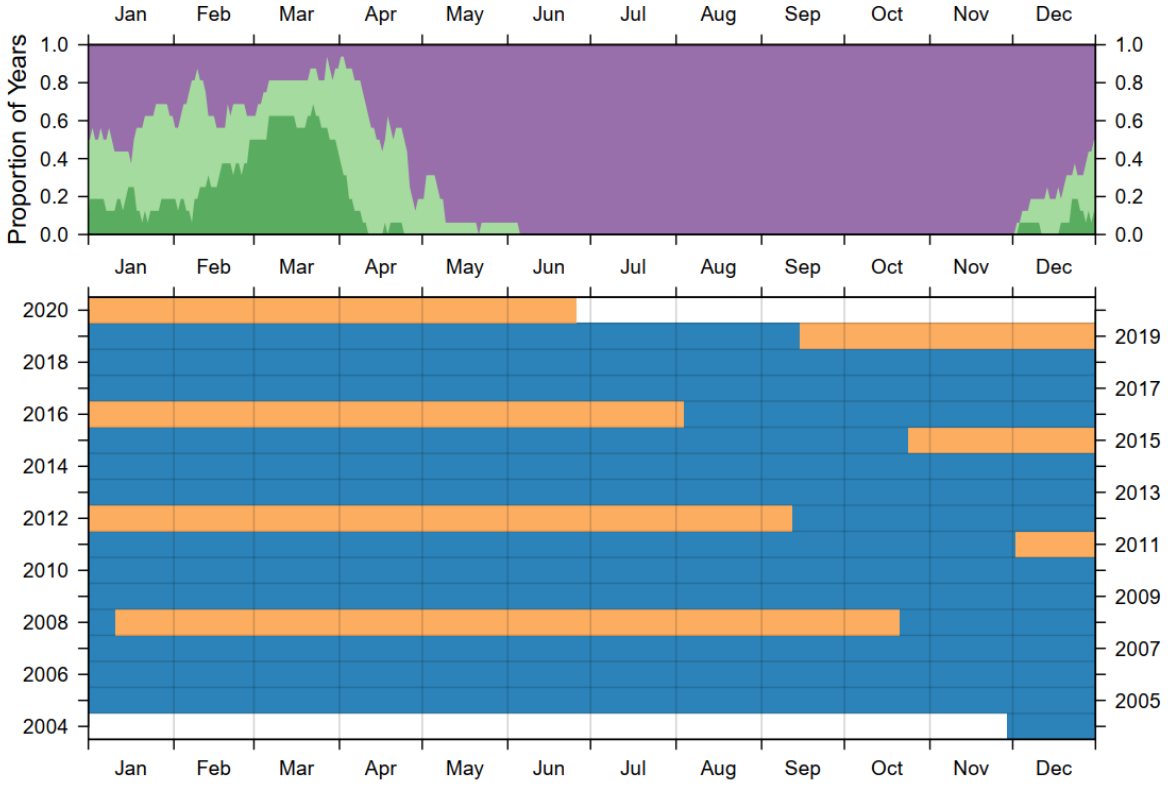

Figure 3: (Top) Proportion of years across the study period (November 29, 2004 through June 25, 2020) with winter-run Chinook salmon absent (purple), detected at low presence ($\leqslant$4.29 daily expanded salvage; light green), or detected at high presence (>4.29 daily expanded salvage; dark green) on a given day of year. (Bottom) Calendar of time slice structure given four slices of 1,422 days per slice with an 80%/20% training (blue)/testing (orange) data split.

Is there a github site where the data and code are stored?

No. The data are all publicly available and across numerous websites and published datasets, but the compiled data and the code are not stored on GitHub or any public repository.

Why did the model perform more poorly when applied to the genetically determined Winter run Chinook?

The model was fit using length-at-date data as genetic data were not compiled and available at the onset of this project in 2021 and did not become compiled until around 2023. Therefore, we would not expect the model to perfectly predict genetic salvage. That said, I am very pleased with how the model performed to predict genetic salvage. Figure 2 in the provided document shows that in 2025 the model successfully predicted genetic absence on 78% of days, successfully predicted genetic low presence on 68% of days, and successfully predicted genetic high presence on 69% of days. Furthermore, we have recommended end users look at the trajectories of the outcome probabilities not just at the daily classification. That is, are we in a window in which the probability of absence is sharply decreasing or are we in a period in which the probability of absence is steady at a probability of nearly 100%.

How specifically is the model used in management? How would you recommend that it be used?

Please see the other document.

Was the logistic regression model presented for illustrative purposes, or is it used in management? How?

The logistic regression model was presented for illustrative purposes to highlight 1) the potential for real time data to predict salvage and 2) to highlight the relationship between two features in the machine learning model and salvage.

Is the Tillotson et al. model also used or was it abandoned? If it was abandoned, why was it?

This question is more suitably addressed by DWR and USBR.