SEAK Pink Salmon Forecast

Sara Miller

Objective

To forecast the Southeast Alaska (SEAK) pink salmon commercial harvest in 2026. This document is for guidance as to what was done for the current forecast year. It is for internal use only.

Executive Summary

Forecasts were developed using an approach originally described in Wertheimer et al. (2006), and modified in Orsi et al. (2016) and Murphy et al. (2019), but assuming a log-normal error structure (Miller et al. 2022). This approach is based on a multiple regression model with the raw juvenile pink salmon catch-per-unit-effort (adj_raw_pink; a proxy for abundance), a vessel factor to account for the survey vessels through time, an odd and even year factor to account for potential odd and even year cycles of abundance, and temperature data from the Southeast Alaska Coastal Monitoring Survey (SECM; Piston et al. 2021; ISTI20_MJJ) or from satellite sea surface temperature (SST) data (Huang et al. 2017). The adj_raw_pink variable is not the same as the CPUE variable used in prior years that was adjusted by the pooled-species vessel calibration coefficient for the Cobb and was the maximum average from either June or July (whichever was higher). Instead, the CPUE term used in the 2025 forecast, the adj_raw_pink variable, is the natural logarithm of the maximum untransformed catch, adjusted to a 20 minute haul, from either June or July. The Stellar and Chellissa vessels were only used for one year each, 2008 and 2009, respectively, and so these two years are not used in the assessment (Table 1).

There were 36 individual models considered:

- adj_raw_pink model with a vessel interaction, an odd/even year factor, and a vessel factor (m1);
- adj_raw_pink model with a vessel interaction, an odd/even year factor, a vessel factor, and temperature data from the the SECM survey (m2);

- 16 adj_raw_pink models with a vessel interaction, an odd/even year factor, a vessel factor, and satellite SST data (m3-m18);
- adjusted CPUE, and an odd/even year factor (m1a);
- adjusted CPUE, an odd/even year factor, and temperature data from the the SECM survey (m2a);
- 16 adjusted CPUE models with an odd/even year factor, and satellite SST data (m3a-m18a);

The model performance metrics one-step ahead mean absolute percent error (MAPE) for the last five years (MAPE5; forecast years 2021 through 2025) was used to evaluate the forecast accuracy of the 36 individual models, the AICc values were calculated for each model to prevent over-parameterization of the model, and the adjusted R-squared values, significant terms, and overall p-value of the model were used to the determine fit. Based upon the performance metric the 5-year MAPE, the AICc values, significant parameters in the models, and the adjusted R-squared values, model xx (a model that included xx; Appendix B) was the best performing model and the 2026-forecast using this model has a point estimate of xx million fish (80% prediction interval: xx to xx million fish).

Analysis

Individual, multiple linear regression models

Biophysical variables based on data from Southeast Alaska were used to forecast the harvest of adult pink salmon in Southeast Alaska, one year in advance, using individual, multiple linear regression models (models m1-m18 and models m1a–m18a). The two simplest regression models (model m1 and model m1a) are as follows. The first model, model m1, consisted of the predictor variable juvenile adj_raw_pink (CPUE) with a vessel factor interaction, an odd/even year factor, and a vessel factor (model m1),

$$E(Y) = \beta_0 + \beta_1 X_{V1} + \beta_2 X_{V2} + \beta_3 X_O + \beta_4 CPUE + \beta_5 (X_{V1} \times CPUE) + \beta_6 (X_{V2} \times CPUE)) + \epsilon.$$

The reference category is an even year with the survey vessel Cobb. The vessel factor and even/odd year factor adjust the intercept, and the vessel interaction terms adjust the model slope. For example, during even years when the survey vessel is the Cobb, the model simplifies to $E(Y) = \beta_0 + \beta_4 CPUE + \epsilon$. Although the simplest model does not contain a temperature variable, including temperature data with CPUE has been shown to result in a substantial improvement in the accuracy of model predictions (Murphy et al. 2019). The temperature index for models m2-m18 was either the SECM survey Icy Strait temperature Index (ISTI20_MJJ; Murphy et al. 2019) or one of the 16 satellite-derived SST data (Huang et al. 2017).

The second model, model m1a, consisted of the predictor variable adjusted CPUE, and an odd/even year factor,

$$E(Y) = \beta_0 + \beta_2 X_O + \beta_3 (CPUE_{adj}) + \epsilon.$$

The reference category (i.e., reference intercept) is an even year. Although the simplest model does not contain a temperature variable, including temperature data with CPUE has been shown to result in a substantial improvement in the accuracy of model predictions (Murphy et al. 2019). The temperature index for models m2a-m18a was either the SECM survey Icy Strait temperature Index (ISTI20_MJJ; Murphy et al. 2019) or one of the 16 satellite-derived SST data (Huang et al. 2017).

The response variable (Y; Southeast Alaska adult pink salmon harvest in millions), the adj_raw_pink, and the adjusted CPUE were log transformed in the model, but temperature data were not. The forecast ($\hat{Y}_i$), and 80% prediction intervals (based on output from program R; R Core Team 2023) from the 36 regression models were exponentiated and bias-corrected (Miller 1984),

$$\hat{F}_i = \exp(\hat{Y}_i + \frac{{\sigma_i}^2}{2}), \tag{2}$$

where $\hat{F}_i$ is the preseason forecast (for each model i) in millions of fish, and σ_i is the variance (for each model i).

Performance metric: One-step ahead MAPE

The model summary results using the performance metric one-step ahead MAPE are shown in Table 2; the smallest value is the preferred model (Appendix C). The performance metric one-step ahead MAPE was calculated as follows.

1. Estimate the regression parameters at time t-1 from data up to time t-1.
2. Make a prediction of $\hat{Y}_t$ at time t based on the predictor variables at time t and the estimate of the regression parameters at time t-1 (i.e., the fitted regression equation).
3. Calculate the MAPE based on the prediction of $\hat{Y}_t$ at time t and the observed value of Y_t at time t,

$$\text{MAPE} = \left|\frac{\exp(Y_t) - \exp(\hat{Y}_t + \frac{{\sigma_t}^2}{2})}{\exp(Y_t)}\right|. \tag{3}$$

4. For each individual model, average the MAPEs calculated from the forecasts,

$$\frac{1}{n}\sum_{t=1}^{n}|\frac{exp(Y_t) - \exp(\hat{Y}_t + \frac{{\sigma_t}^2}{2})}{\exp(Y_t)}|, \tag{4}$$

   where n is the number of forecasts in the average (5 forecasts for the 5-year MAPE and 10 forecasts for the 10-year MAPE). For example, to calculate the five year one-step-ahead MAPE for model m1 for the 2022 forecast, use data up through year 2016 (e.g., data up through year 2016 is t -1 and the forecast is for t, or year 2017). Then, calculate a MAPE based on the 2017 forecast and the observed pink salmon harvest in 2017 using equation 3. Next, use data up through year 2017 (e.g., data up through year 2017 is t -1 and the forecast is for year 2018; t) and calculate a MAPE based on the 2018 forecast and the observed pink salmon harvest in 2018 using equation 3. Repeat this process for each subsequent year through year 2020 to forecast 2021. Finally, average the five MAPEs to calculate a five year one-step-ahead MAPE for model m1. As the results of the 5-year MAPEs with or without the forecast bias adjustment have been similar (i.e., the model performance order did not change whether the five year one-step-ahead MAPE or the bias-corrected five year one-step-ahead MAPE was compared), for simplicity, the bias adjustment for the forecast was not used in the calculation of the five year one-step-ahead MAPE for model comparison.

Akaike Information Criterion corrected for small sample sizes (AICc)

Hierarchical models were compared with the AICc criterion. The best fit models, according to the AICc criterion, is one that explains the greatest amount of variation with the fewest independent variables (i.e., the most parsimonious; Table 2). The lower AICc values are better, and the AICc criterion penalizes models that use more parameters. Comparing the AICc values of two hierarchical models, a $\Delta_i \leq 2$ suggests that the two models are essentially the same, and the most parsimonious model should be chosen (Burnham and Anderson 2004). If the $\Delta_i > 2$, the model with the lower AICc value should be chosen.

Results