Mathematical Model Descriptions of the Inseason Projection for Canadian-origin Yukon Chinook Salmon

Aaron Lambert

3/26/2024

JTC Preseason Forecast

The initial value for the estimated run size in log space in year y, $\ln(\widehat{R}_y)$, was drawn from a normal distribution with a mean equal to the natural log of the JTC preseason forecast point estimate μ_y and a standard deviation σ_μ equal to the empirical standard deviation of lognormal prediction errors ϵ_y of the JTC preseason forecast over the years $(N = 16)$ during which the current preseason forecast methodology was used (2007-2022):

(1)
$$\ln(\widehat{R}_y) \sim \mathrm{Normal}(\ln(\hat{\mu}_y), \sigma_\mu)$$

where the empirical standard deviation was calculated as

(2)
$$\sigma_\mu = \sqrt{\frac{\sum_y^N (\epsilon_y - \bar{\epsilon})^2}{N - 1}}.$$

Annual preseason forecast errors were calculated as log ratio of the preseason forecast predicted run size and the observed run size as:

(3)
$$\epsilon_y = \ln\left(\frac{\widehat{R}_y^{\text{preseason}}}{R_y}\right).$$

The initial run size projection $\widehat{R}_y$, drawn from the preseason forecast prior, was then updated based on the relationship between cumulative daily PSS Chinook salmon passage (i.e., the sum of daily Chinook passage at PSS through the current point in the season) and the run size R_y.

Pilot Station Sonar (PSS) Likelihood

Three alternative methods were explored to generate run size predictions based on PSS passage observations through a given point in the season and update the projected run size from the preseason forecast prior.

Method 1: Regressing Observed Cumulative PSS Passage Against Run Size (Model *PSSreg*)

For the first method, three different likelihood functions were constructed and evaluated. In model *PSSreg*, the linear relationship between annual run size and cumulative PSS passage in previous years was estimated as:

$$\hat{R}_y^{PSS} = \alpha_D + \beta_D \sum_{d=148}^{D} PSS_{d,y} \quad (4)$$

where α_D and β_D are the intercept and slope coefficients, respectively, estimated using data up to day of year D (the day the inseason projection is made), and $PSS_{y,d}$ was the PSS passage on a given day d in year y. Indexing begins on day of year 148 (May 28) because this was the earliest observation of salmon passage at PSS across years of project operation. Slope and intercept parameters were estimated by minimizing the difference between the natural log of projected run size $\hat{R}_y^{PSS}$ and that of the true but observed run size R_y across years 2005-2022 where,

$$\ln\left(R_y\right) \sim Normal(\ln\left(\hat{R}_y^{PSS}\right), \sigma_D^{PSS}) \quad (5)$$

and σ_D^{PSS} was the residual log-normal standard deviation for predictions based on PSS passage through day D and were directly estimated. Priors on regression parameters were uninformative and bounded at zero as,

$$\begin{gathered} \alpha_D \sim Normal(0, 10^{15})[0, \infty) \\ \beta_D \sim Normal(0, 10^{15})[0, \infty) \\ \sigma_D^{PSS} \sim Normal(0,10)[0, \infty). \end{gathered} \quad (6)$$

The current estimate of run size $\hat{R}_y$ was updated from the preseason forecast with the likelihood

$$\ln(\hat{R}_y) \sim Normal\left(\ln(\hat{R}_y^{PSS}), \sigma_D^{PSS^*}\right) \quad (7)$$

where $\sigma_D^{PSS^*}$ was the empirical standard deviation for run size predictions based on PSS passage estimates through day D for years 2007-2022, and was calculated as

(8)

$$\sigma_D^{PSS^*} = \sqrt{\frac{\sum_{y=2007}^{y=2022}(\epsilon_y^{PSS} - \bar{\epsilon}^{PSS})}{N-1}}.$$

The PSS projection errors ϵ_y^{PSS} were calculated as the difference between the observed and PSS projected run sizes in log space as

(9)

$$\epsilon_y^{PSS} = \ln\left(\frac{\hat{R}_y^{PSS}}{R_y}\right).$$

Comparison of projection errors were restricted to years 2007-2022 to ensure that the projection based on PSS salmon passage were appropriately weighted and comparable to the uncertainty associated with the preseason forecast prediction when updating the joint posterior for run size $\hat{R}_y$.

Method 2: Fitting Yearly Arrival Distributions to Estimate Total PSS Passage (Models *PSSnormal_ESprop*, *PSSlogistic_ESprop*)

The relationship between cumulative PSS passage and run size becomes more informative as the season progresses (Figure 2). The second method of relating PSS Chinook salmon passage to run size included two variations of fitting either a normal distribution (Model *PSSnormal_ESprop*) or a logistic function (Model *PSSlogistic_ESprop*) to observed daily or cumulative PSS passage, respectively, and using these distributions to generate estimates of the total PSS Chinook salmon passage at the end of the season. Given the known challenge in fitting either of these distributions to mid-season PSS passage (uncertainty associated with midpoint, spread and magnitude of the run), informative empirical prior distributions were specified for the parameters of these distributions based on the shape and scale daily, or cumulative, of PSS passage distributions in previous years. After predicting PSS passage for the remaining portion of the season using the fitted distribution, the predicted total PSS passage was then used in a regression relationship to predict the run size for the year of interest and to update the preseason forecast as before (see Eq.(7)).

To develop informative empirical prior distributions for the shape and scale parameters describing the distribution of Chinook salmon passing PSS to inform inseason fitting, the normal and logistic curves were fit to complete PSS passage data from previous years (2005–2022) by minimizing the sum of squares external to the Stan model using the BBMLE package (Bolker and R Development Core Team 2022). This external fitting procedure resulted in vectors of parameter estimates for the shape and scale of salmon passage distributions in each previous year (2005–2022). These parameter estimates were then used to construct informative priors within the Stan model when fitting a passage distribution to partial PSS count data up to the current day D in the season.

Next, to quantify the accuracy of predictions for run size based on projected total PSS passage in previous years, normal curves were fit to observed data for current and historical years (2005 – 2022) up to day D. In the case of fitting a normal distribution to daily passage for model *PSSnormal_ESprop* a curve was fit as

$$\mathrm{PSS}^{\mathrm{Obs}}_{\mathrm{y,d=[148,D]}} = \alpha_{\mathrm{y}} \left(\frac{1}{\sigma_{\mathrm{y}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mathrm{d}-\mu_{\mathrm{y}}}{\sigma_{\mathrm{y}}}\right)^2} \right) \quad (10)$$

where α_{y} was the scale parameter, μ_{y} was the midpoint (i.e., mean of the normal distribution), and σ_{y} was the standard deviation of the normal distribution in year y. Informative priors were placed on parameters as informed by the shape and scale of PSS passage distributions in previous years as

$$\mu_{\mathrm{y}} \sim \mathrm{Normal}\left(\mathrm{mean}(\boldsymbol{\mu}^{MLE}), \mathrm{sd}(\boldsymbol{\mu}^{MLE})\right)[165,181]$$

(11) $$\sigma_y \sim \mathrm{Normal}\left(\mathrm{mean}(\boldsymbol{\sigma}^{\mathrm{MLE}}), \mathrm{sd}(\boldsymbol{\sigma}^{\mathrm{MLE}})\right)[1,14]$$

$$\alpha_y \sim \mathrm{Normal}\left(\mathrm{mean}(\boldsymbol{\alpha}^{MLE}), \mathrm{sd}(\boldsymbol{\alpha}^{MLE})\right)[10^4, 4\mathrm{x}10^5].$$

Where $\boldsymbol{\mu}^{MLE}$, $\boldsymbol{\sigma}^{\mathrm{MLE}}$, and $\boldsymbol{\alpha}^{MLE}$ are the vectors of the year-specific (2005-2022) passage distribution parameters estimated prior to model fitting. Bounds on parameters were chosen to constrain estimates to plausible values for midpoint, spread and scale of PSS passage.

Next, a prediction for the unobserved remainder of the PSS daily passage was made for the current and historical years y from day $\mathrm{D}+1$ to the final day of the season, D^*as

(12) $$\widehat{\mathrm{PSS}}^{\mathrm{Pred}}_{\mathrm{d}=[\mathrm{D}+1,\mathrm{D}^*],\mathrm{y}} = \alpha_y\left(\frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mathrm{d}-\mu_y}{\sigma_y}\right)^2}\right).$$

The predicted total PSS passage in each year was then calculated as the sum observed PSS passage up to the forecast day D, plus the estimated passage for the remainder of the season based on Eq.(12).

Normal distribution parameters describing daily PSS passage were estimated by minimizing the difference between predicted PSS passage $\widehat{PSS}^{Pred}_{\mathrm{d,y}}$ and the true, observed PSS passage $PSS^{Obs}_{d,y}$ for values greater than zero where,

(13) $$\ln\left(\mathrm{PSS}^{Obs}_{\mathrm{d}=[148,\mathrm{D}],\mathrm{y}}\right) \sim \mathrm{Normal}(\ln\left(\widehat{PSS}^{Pred}_{\mathrm{d}=[148,\mathrm{D}],\mathrm{y}}\right), \sigma^{\mathrm{PSS}}_{\mathrm{y}})$$

and $\sigma^{\mathrm{PSS}}_{\mathrm{y}}$ was directly estimated from the data. Priors on $\sigma^{\mathrm{PSS}}_{\mathrm{y}}$ were mildly informative in log-space and drawn from a half normal as

(14) $$\ln\left(\sigma^{\mathrm{PSS}}_{\mathrm{y}}\right) \sim \mathrm{Normal}(0,5)[0,\infty).$$

The complete PSS passage for each year $\widehat{TPSS}_y$ was then calculated by adding the predicted remaining daily passage to the observed daily passage through day D as

(15) $$\widehat{\mathrm{TPSS}}_{\mathrm{y}} = \Sigma^{\mathrm{D}}_{\mathrm{d}=148}\mathrm{PSS}^{\mathrm{Obs}}_{\mathrm{d,y}} + \Sigma^{\mathrm{D}^*}_{\mathrm{d}=\mathrm{D}+1}\widehat{\mathrm{PSS}}^{\mathrm{Pred}}_{\mathrm{d,y}}.$$

A regression on run size was fit based on a linear relationship between historic total cumulative PSS Chinook passage through the final day of the season D^*, as

(16)

$$\widehat{\mathrm{R}}_{\mathrm{y}}^{\mathrm{PSS}} = \alpha_{\mathrm{D}^*} + \beta_{\mathrm{D}^*} \sum_{\mathrm{d}=148}^{\mathrm{D}^*} \mathrm{PSS}_{\mathrm{d,y}}.$$

As before, slope and intercept parameters were estimated by minimizing the difference between the projected run size $\widehat{\mathrm{R}}_{\mathrm{y}}^{\mathrm{PSS}}$ and the true, observed run size R_{y} as in equation (5). Priors on regression parameters remained uninformative as in equation (6). The resulting regression relationship was then used to predict the run size in the forecast year using the total PSS estimate $\widehat{TPSS}_y$ and the preseason forecast was then updated as before in equations (7), (8), and (9), conditional on the estimates of α_{D^*} and β_{D^*}.

In model *PSSlogistic_ESprop*, a logistic function was used to represent cumulative daily PSS passage up to the current day in the season D as

(17)

$$CPSS_{D,y}^{Pred} = \mathrm{A}_{\mathrm{y}}\left(\frac{1}{1+\mathrm{e}^{-\frac{\mathrm{d}-\mathrm{M}_{\mathrm{y}}}{\mathrm{S}_{\mathrm{y}}}}}\right)$$

where A_{y} is the scale parameter, M_{y} is the midpoint (i.e., mean of the logistic function), and S_{y} is the parameter describing the steepness of the logistic function at the inflection point. Informative priors were placed on parameters as informed by estimates of logistic parameters fit to cumulative PSS passage distributions in previous years and bounded at zero as

(18)

$$\mathrm{M}_{\mathrm{y}} \sim \mathrm{Normal}\left(\mathrm{mean}(\mathbf{M}^{MLE}), \mathrm{sd}(\mathbf{M}^{MLE})\right)[0,\infty)$$

$$S_{\mathrm{y}} \sim \mathrm{Normal}\left(\mathrm{mean}(\boldsymbol{S}^{MLE}), \mathrm{sd}(\mathbf{S}^{MLE})\right)[0,\infty)$$

$$\mathrm{A}_{\mathrm{y}} \sim \mathrm{Normal}\left(\mathrm{mean}(\mathbf{A}^{MLE}), \mathrm{sd}(\mathbf{A}^{\mathrm{MLE}})\right)[0,\infty),$$

where $\mathbf{M}^{\mathrm{MLE}}$, $\mathbf{S}^{\mathrm{MLE}}$, and $\boldsymbol{\alpha}^{MLE}$ are vectors of year-specific logistic parameters describing the cumulative PSS passage distribution in previous years (2005-2022). This process was similar to the procedure in Model *PSSnormal_ESprop*, with the difference being a logistic function as opposed to a normal distribution being fit to cumulative rather than daily PSS passage data.

Logistic function parameters describing cumulative PSS passage were estimated by minimizing the difference between projected cumulative run size $\sum_{d=148}^{D^*} \widehat{\mathrm{R}}_{\mathrm{y}}^{\mathrm{PSS}}$ and the true, observed cumulative run size $\sum_{d=148}^{D^*} R_y$ for values greater than zero where,

$$\ln\left(CPSS_{D,y}^{Obs}\right) \sim Normal\left(\ln\left(\widehat{CPSS}_{D,y}^{Pred}\right), \sigma_y^{CPSS}\right) \quad (19)$$

and $\sigma_{\mathrm{y}}^{\mathrm{CPSS}}$ was directly estimated. Priors on $\sigma_{\mathrm{y}}^{\mathrm{CPSS}}$ were mildly informative in log-space and drawn from a half normal as

$$\sigma_{\mathrm{y}}^{\mathrm{CPSS}} \sim \mathrm{Normal}(0{,}5)[0, \infty). \quad (20)$$

The complete PSS passage $\widehat{TPSS}_y$ for each year was then assumed to be equal to the scaling parameter A_y as the scaling parameter represents the total estimated PSS Chinook salmon passage. A regression on run size was fit based on a linear relationship between historic total cumulative PSS Chinook passage through the final day of the season D^* as seen in equation **Error! Reference source not found.**) and the resulting regression relationship was used to predict the run size in the forecast year. The preseason forecast was then updated as before in equations (7), (8), and (9), conditional on the estimates of α_{D^*} and β_{D^*}.

Method 3: Estimating Mean Daily Total PSS Proportion to Estimate total PSS Passage (Model: *PSSprop_ESprop*)

Like method two, method three leverages the improved relationship between total PSS passage and run size (Figure 2), as opposed to midseason cumulative PSS passage. In method three, one logistic function was fit to the observed historic proportion of PSS Chinook salmon passage for years 2005-2022, representing the mean expected proportion of total PSS passage observed on each day of the season. The expected proportion was then used to obtain a midseason estimate of the total PSS Chinook salmon passage. The predicted total PSS passage was then used in a regression relationship to predict the run size for the year of interest and update the preseason forecast.

First, a single logistic function was fit to the observed historic proportion of cumulative PSS passage as

(21) $$\hat{P}_d^{PSS} = \frac{1}{1 + e^{-\frac{d-m}{s}}}$$

where m was the midpoint parameter, s was the steepness parameter, and $\hat{P}_d^{PSS}$ was the expected proportion of Chinook salmon passing PSS on day d. Uniform priors with informative lower and upper limits were assumed for logistic model parameters:

(22) $$\mathrm{m} \sim \mathrm{Uniform}(166{,}182)$$
$$s \sim Uniform(1{,}12).$$

Parameters were estimated by minimizing the difference between the projected cumulative proportion of Chinook salmon passing PSS and the true, observed proportion $P_{d,y}^{PSS}$ where,

(23) $$\mathrm{P}_{\mathrm{d,y}}^{PSS} \sim Normal\left(\hat{P}_d^{PSS}, \sigma_P\right)$$

and σ_P was directly estimated. The priors for σ_P were uninformative:

(24) $$\sigma_{\mathrm{P}} \sim Normal(0{,}100).$$

The total PSS passage in each year y was then estimated by dividing the observed cumulative PSS passage in each year y through day of season D, by the predicted average proportion on that day $\hat{P}_d^{PSS}$ as

(25) $$\widehat{TPSS}_y = \frac{\sum_{d=148}^{D} PSS_y^{obs}}{\hat{P}_D^{PSS}}$$

and the run size was then estimated using the EOS regression as in equation (25). As before, slope and intercept parameters were estimated by minimizing the difference between the projected run size $\widehat{\mathrm{R}}_{\mathrm{y,D}}^{\mathrm{PSS}}$ and the true, observed run size R_{y} in each year (2005 – 2022) as in equation (5). Priors on regression7 parameters remained uninformative as in equation (6). The joint posterior for run size estimate for the current year was then updated from the preseason forecast prior as before using equations (7), (8), and (9) based on the likelihood of the inseason PSS-based estimate of run size.

Eagle Sonar Prediction (Models: *PSSreg_ESreg*, *PSSreg_ESrefIF*, *PSSreg_ESprop*, *PSSnormal_ESprop*, *PSSlogistic_ESprop*, *PSSprop_ESprop*)

After Canadian-origin Chinook salmon pass the PSS site, they travel approximately 1770 km to the US-Canadian border where daily passage is estimated at the Eagle Sonar site (Fig. 1). Like Pilot Station Sonar passage, Eagle Sonar passage provides information on Canadian-origin Chinook salmon run size. The incorporation of Eagle Sonar daily passage information into the projection model (which includes PSS predictions) was explored using three different methods.

For the first two methods (Models *PSSreg_ESreg* and *PSSreg_ESregIF*), a regression between cumulative daily Eagle Sonar passage and the run size was fit as

$$\hat{R}_y^{ES} = \alpha_D + \beta_D \sum_{d=148}^{D} ES_{d,y} \quad (26)$$

where α_D and β_D were the intercept and slope coefficients, respectively, estimated using data up to day D, and $ES_{y,d}$ was the Eagle Sonar passage. Slope and intercept parameters were estimated by minimizing the difference between predicted Canadian-origin run size based on Eagle Sonar passage $\hat{R}_y^{ES}$ and the true, observed Canadian-origin run size R_y in the same manner as equation (5). Priors on regression parameters remained uninformative as in equation (6) and the preseason forecast was then updated as before in equations (7), (8), and (9).

The models *PSSreg_ESreg* and *PSSreg_ESregIF* differed in whether the Eagle Sonar likelihood component of the projection was incorporated into the model from June 1 (the beginning of the projection; *PSSreg_ESreg*) or upon first observed passage of Chinook salmon at

Eagle Sonar in the year the projection is made (*PSSreg_ESregIF*). The difference in these methods resulted from the fact that Chinook salmon do not arrive at Eagle Sonar until the end of June, meaning that Model *PSSreg_ESreg* provided information on run size based only on the intercept parameter before Eagle Sonar abundance estimates were available. This resulted in a projection based on the average historical run size until approximately July 1, whereas only using the Eagle Sonar likelihood to update the joint posterior after the first Chinook salmon passes Eagle Sonar in the projection year (Model *PSSreg_ESregIF*) may be more informative and result in improved accuracy and/or precision because the mean Eagle Sonar run size would not influence the projection until inseason border passage was available in the projection year.

For the third method (Models *PSSreg_ESprop*, *PSSnormal_ESprop*, *PSSlogistic_ESprop*, *PSSprop_ESprop*), a simple proportion estimator was used to project the run size using Eagle Sonar passage information. First, the observed proportion $\mathrm{P}_{\mathrm{y,D}}^{ES}$ of total Chinook salmon passing Eagle Sonar through day D in year y of the total run size R was calculated as

(27)
$$\mathrm{P}_{\mathrm{y,D}}^{\mathrm{ES}} = \frac{\sum_{\mathrm{d}}^{\mathrm{D}} \mathrm{ES}_{\mathrm{y,d}}}{\mathrm{R}_{\mathrm{y}}}.$$

Next, the mean proportion across years (2005 – 2022) observed was calculated as

(28)
$$\bar{P} = \frac{\sum_{y=2005}^{y=2022} P_{y,D=[d,D]}^{ES}}{N_y}$$

where N_y was the number of years with observed Eagle Sonar data. and an Eagle Sonar projection for the run size was calculated as

(29)
$$\widehat{\mathrm{R}}_{\mathrm{y}}^{ES} = \frac{\sum_{\mathrm{d}}^{\mathrm{D}} \mathrm{ES}_{\mathrm{y,d}} + 1}{\bar{P}}.$$

The joint posterior probability for the projected run size $\hat{R}_y$ was then updated based on the estimate derived from the current Eagle Sonar passage using $\hat{R}_y^{ES}$ as before (Eq. (7), (8), and (9)).

Estimating PSS Canadian-origin Proportion of Chinook Salmon (Model *PSSreg_GSI*)

Genetic stock identification data describe the proportion of total Chinook salmon passing PSS that are assigned to the Canadian-origin stock via genetic sampling. It was hypothesized that Canadian-origin run size might be better predicted by the cumulative PSS passage of Canadian-origin Chinook salmon through a given day in the season D, rather than total passage as previously described, as:

(30)
$$\hat{R}_y^{PSS} = \alpha_D + \beta_D \sum_{d}^{D} \left(\hat{g}_d * PSS_{y,d}\right)$$

where $\hat{g}_d$ is the estimated proportion of Canadian-origin Chinook salmon passing PSS on day d and $PSS_{y,d}$ is the observed total number of Chinook salmon passing PSS on day d in year y.

ADF&G aggregate GSI samples into three temporal strata and were therefore not available on a daily basis. To estimate the proportion of Canadian-origin Chinook salmon at PSS for each day of the season as needed for daily run size estimation, we represented the genetic stock proportion as a smooth transition throughout the season from a high proportion of Canadian-origin Chinook salmon at the start of the run, to a relatively low proportion at the end of the run. A random walk representing the daily GSI proportion of Canadian-origin Chinook salmon was estimated with an inverse-logit transformation to scale expected Canadian-origin proportions onto the appropriate [0,1] range. To begin, an initial value for the proportion on day $d = 148$ was drawn from a normal distribution as

(31)
$$\hat{g}_{d=148}^{logit} \sim Normal(0,1)$$

and following days up to the day of interest D were drawn as

(32) $$\hat{g}_{d}^{logit} \sim \mathrm{Normal}(\hat{g}_{d-1}^{logit}, \phi)$$

where ϕ was the estimated standard deviation for the normal random walk with an informative prior specified as:

(33) $$\ln(\phi) \sim \mathrm{Normal}(-1,1).$$

Next, to re-scale daily predicted proportions from the normal random walk on the appropriate 0 – 1 range, predicted values were inverse logit transformed as

(34) $$\hat{g}_{d} = \frac{e^{\hat{g}_{d}^{logit}}}{1 + e^{\hat{g}_{d}^{logit}}}$$

and the likelihood for the rescaled predicted Canadian proportion on day d ($\hat{g}_d$), given the GSI data, was calculated from a beta distribution as

(35) $$\hat{g}_{d} \sim \mathrm{Beta}\left(A_{d}^{prior}, B_{d}^{prior}\right)$$

where the shape parameters are reparametrized in terms of the mean and standard deviation as follows:

(36) $$A_{y,d}^{prior} = \left(\frac{1-\bar{g}_{d}}{\bar{\tau}_{d}^{2}} - \frac{1}{\bar{g}_{d}}\right) * \bar{g}_{d}^{2}$$

$$B_{d}^{prior} = A_{d}^{prior} * \left(\frac{1}{\bar{g}_{d}} - 1\right).$$

The expected proportion $\bar{g}_d$ and standard deviation $\bar{\tau}_d$ were calculated from the observed historic PSS Canadian-origin passage proportions on day d across years y as

(37) $$\bar{g}_{d} = \frac{\sum_{y}^{Y} g_{d,y}}{N_{y}}$$

$$\bar{\tau}_{d} = \sqrt{\frac{\sum_{y}^{Y}(g_{d,y} - \bar{g}_{d})}{N_{y}}}$$

where N_y is the number of years where a GSI sampling strata encompassed day d.