Directions for Inseason Estimates for Pink Salmon Cumulative Harvest in Southeast Alaska

Sara Miller

2019-07-16

Summary

The regular inseason forecasts of cumulative traditional seine harvest in Southern and Northern Southeast are made using multiple regression models that relate weekly cumulative harvest and sex ratios to the final harvest. The data is updated weekly by the pink salmon biologist and models are run throughout the season.

Data Setup

Prior to the season, update the two data files in the folder data/ with the most recent cumulative catch (millions) by statistical week and sex percetage (i.e. the percentage of males from the traditional purse seine harvest). The Northern Southeast (NSE) model uses harvest data from 1984 to 2018, stat weeks 27-35, and sex ratios from District 12; the Southern Southeast (SSE) model uses harvest data from 1983 to 2018, stat weeks 28-35, and sex ratios from District 4. This information can be downloaded from…

For the current year, duplicate the data for the prior year until information is available.

Inseason, these data files will be updated weekly to run weekly estimates of cumulative purse sein pink salmon harvest in Districts 9-14 (NSE) and Districts 1-7 (SSE).

Analysis

All required packages need for the analysis are loaded prior to running models. These names of these pacakges are located at the top of the *function.r* file. If you are missing a package, you will need to install it from *CRAN* in the usual fashion.

The functions are sourced within the *nse_analysis.r file* and *sse_analysis.r file* from the file *functions.r file.* The functions are:

sex_dev = creates sex ratio deviation from mean from raw data

nll = sets up structure of negative log likelihood linear models

nll_sex = sets up structure of negative log likelihood linear models that include the parameter sex

f_est = estimates the model fits

f_params = outputs the intercept, slope, and sigma for all model fits

f_pred = outputs the parameters and predictions for all model fits

model_select = outputs the AIC, AICc, delta AICc, AICc weight, and model averaged prediction by week

The top models are then presented.

Inseason Estimates

The inputs that need to be specified for the code to run are the stat week of the analysis and the year of the anlaysis. These are specified at the top of the *nse_analysis.r* and the *sse_analysis.r* code in the code fragments:

```
# Weekly Settings
sw_forecast <- 29   # forecast week
year_forecast <- 2019   # forecast year
```

Model

The general inseason forecasting model can be written as,

$$H_t = (a_w + b_w * E_{t,w} + c_w * H_{t,w}) * (1 + d_w * S_{t,w})$$

where H_t is total harvest in year t;
$E_{t,w}$ is cumulative CPUE by the end of week w in year t;
H_{tw} is cumulative harvest by the end of week w in year t;
$S_{t,w}$ is the sex ratio index by the end of week w in year t;
a_w, b_w, c_w, and d_w are parameters.

The sex ratio index is derived from deviations of weekly male proportions to the corresponding weekly average. It is the value of cumulative deviations over time in a particular year.

In the preliminary study, $E_{t,w}$ was found to be a redundant predictor. After removing $E_{t,w}$ (let $b_w = 0$) from the model above, the reduced model is,

$$H_t = (a_w + c_w * H_{t,w}) * (1 + d_w * S_{t,w}),$$

Further, if $d_w = 0$, a more parsimonious model is

$$H_t = a_w + c_w * H_{t,w}.$$

Model Selection

Akaike Information Criterion (AIC) or AIC corrected for small sample sizes (AICc values; Burnham and Anderson 1998) is used for model selection. The better model has the smallest AIC or AICc value.The difference ($\triangle_i$) between a given model and the model with the lowest AICc value is the primary statistic for choosing appropriate models. For biologically realistic models, those with $\triangle_i \leq 2$ have substantial support, those in which $4 \leq \triangle_i \leq 7$ have considerably less support, and models with $\triangle_i > 10$ have essentially no support (Burnham and Anderson 2004).

Model Averaging

To determine the plausibility of each model, given the data and set of models, the 'Akaike weight' w_i of each model was calculated,

$$w_i = \frac{exp(-\triangle_i/2)}{\sum_{r=1}^{R} exp(-\triangle_r/2)}.$$

From the available set of models, the 'Akaike weight' is considered as the weight of evidence in favor of a particular model i being the actual best model (Akaike 1983; Burnham and Anderson 2002). The 'average' model was determined by averaging the predicted response variable Y across the avilable set of models, and then using the corresponding w_i's as weights. The model averaged prediction is then

$$\bar{Y} = \sum_{i=1}^{I} w_i * \hat{Y}_i.$$

The inseason forecast will be produced in the section:

```
# top models
results %>%
  filter(delta_AICc==0)
```

The forecast is in the column 'model_avg.'

References

Akaike, H., 1983. Information measures and model selection. International Statistical Institute 44, 277-291.

Burnham, K. P., and Anderson, D.R. 1998. Model Selection and Inference. Springer, New York. 353 pp.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York.

Burnham, K. P., and Anderson, D.R. 2004. Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods and Research, Vol. 33(2): 261-304